United States Patent Application

CAMPYLOBACTER GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF GANGLIOSIDES AND GANGLIOSIDE MIMICS

Inventors: Michel Gilbert, a Canadian citizen, residing at 116-101 Sacré-Coeur, Hull, Quebec CANADA   J8X 1C7

Warren W. Wakarchuk, a Canadian citizen, residing at #11-837 Eastvale Drive, Gloucester, Ontario CANADA   K1J 7T5

Assignee: National Research Council of Canada

CAMPYLOBACTER GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF GANGLIOSIDES AND GANGLIOSIDE MIMICS

CROSS-REFERENCE TO RELATED APPLICATIONS

5

This application claims benefit of US Provisional Application No. 60/118,213, which was filed on February 1, 1999, and is a continuation-in-part of US Application No. 09/495,406 filed January 31, 2000, both of which are incorporated herein by reference for all purposes.

10

BACKGROUND OF THE INVENTION

Field of the Invention

This invention pertains to the field of enzymatic synthesis of oligosaccharides, including gangliosides and ganglioside mimics.

Background

15

Gangliosides are a class of glycolipids, often found in cell membranes, that consist of three elements. One or more sialic acid residues are attached to an oligosaccharide or carbohydrate core moiety, which in turn is attached to a hydrophobic lipid (ceramide) structure which generally is embedded in the cell membrane. The ceramide moiety includes a long chain base (LCB) portion and a fatty acid (FA) portion. Gangliosides, as well as other glycolipids and their structures in general, are discussed in, for example, Lehninger,

20

Biochemistry (Worth Publishers, 1981) pp. 287-295 and Devlin, *Textbook of Biochemistry* (Wiley-Liss, 1992). Gangliosides are classified according to the number of monosaccharides in the carbohydrate moiety, as well as the number and location of sialic acid groups present in the carbohydrate moiety. Monosialogangliosides are given the designation "GM", disialogangliosides are designated "GD", trisialogangliosides "GT", and

25

tetrasialogangliosides are designated "GQ". Gangliosides can be classified further depending on the position or positions of the sialic acid residue or residues bound. Further classification is based on the number of saccharides present in the oligosaccharide core, with the subscript "1" designating a ganglioside that has four saccharide residues (Gal-GalNAc-Gal-Glc-

Ceramide), disaccharide (Gal-Glc-Ceramide) and monosaccharide (Gal-Ceramide) gangliosides, respectively.

Gangliosides are most abundant in the brain, particularly in nerve endings. They are believed to be present at receptor sites for neurotransmitters, including acetylcholine, and can also act as specific receptors for other biological macromolecules, including interferon, hormones, viruses, bacterial toxins, and the like. Gangliosides are have been used for treatment of nervous system disorders, including cerebral ischemic strokes. *See, e.g.*, Mahadnik *et al.* (1988) *Drug Development Res.* 15: 337-360; US Patent Nos. 4,710,490 and 4,347,244; Horowitz (1988) *Adv. Exp. Med. and Biol.* 174: 593-600; Karpiatz *et al.* (1984) *Adv. Exp. Med. and Biol.* 174: 489-497. Certain gangliosides are found on the surface of human hematopoietic cells (Hildebrand *et al.* (1972) *Biochim. Biophys. Acta* 260: 272-278; Macher *et al.* (1981) *J. Biol. Chem.* 256: 1968-1974; Dacremont *et al. Biochim. Biophys. Acta* 424: 315-322; Klock *et al.* (1981) *Blood Cells* 7: 247) which may play a role in the terminal granulocytic differentiation of these cells. Nojiri *et al.* (1988) *J. Biol. Chem.* 263: 7443-7446. These gangliosides, referred to as the "neolacto" series, have neutral core oligosaccharide structures having the formula $[Gal\beta\text{-}(1,4)GlcNAc\beta(1,3)]_n Gal\beta(1,4)Glc$, where n = 1-4. Included among these neolacto series gangliosides are $3'\text{-}nLM_1$ $(NeuAc\alpha(2,3)Gal\beta(1,4)GlcNAc\beta(1,3)Gal\beta(1,4)\text{-}Glc\beta(1,1)\text{-}Ceramide)$ and $6'\text{-}nLM_1$ $(NeuAc\alpha(2,6)Gal\beta(1,4)GlcNAc\beta(1,3)Gal\beta(1,4)\text{-}Glc\beta(1,1)\text{-}Ceramide)$.

Ganglioside "mimics" are associated with some pathogenic organisms. For example, the core oligosaccharides of low-molecular-weight LPS of *Campylobacter jejuni* O:19 strains were shown to exhibit molecular mimicry of gangliosides. Since the late 1970s, *Campylobacter jejuni* has been recognized as an important cause of acute gastroenteritis in humans (Skirrow (1977) *Brit. Med. J.* 2: 9-11). Epidemiological studies have shown that *Campylobacter* infections are more common in developed countries than *Salmonella* infections and they are also an important cause of diarrheal diseases in developing countries (Nachamkin *et al.* (1992) *Campylobacter jejuni: Current Status and Future Trends.* American Society for Microbiology, Washington, DC.). In addition to causing acute gastroenteritis, *C. jejuni* infection has been implicated as a frequent antecedent to the development of Guillain-Barré syndrome, a form of neuropathy that is the most common

2

cause of generalyzed paralysis (Ropper (1992) *N. Engl. J. Med. 326*: 1130-1136). The most common *C. jejuni* serotype associated with Guillain-Barré syndrome is O:19 (Kuroki (1993) *Ann. Neurol.* 33: 243-247) and this prompted detailed study of the lipopolysaccharide (LPS) structure of strains belonging to this serotype (Aspinall *et al.* (1994a) *Infect. Immun.* 62: 2122-2125; Aspinall *et al.* (1994b) *Biochemistry* 33: 241-249; and Aspinall *et al.* (1994c) *Biochemistry* 33: 250-255).

Terminal oligosaccharide moieties identical to those of GD1a, GD3, GM1 and GT1a gangliosides have been found in various *C. jejuni* O:19 strains. *C. jejuni* OH4384 belongs to serotype O:19 and was isolated from a patient who developed the Guillain-Barré syndrome following a bout of diarrhea (Aspinall *et al.* (1994a), *supra.*). It was showed to possess an outer core LPS that mimics the tri-sialylated ganglioside GT1a. Molecular mimicry of host structures by the saccharide portion of LPS is considered to be a virulence factor of various mucosal pathogens which would use this strategy to evade the immune response (Moran *et al.* (1996a) *FEMS Immunol. Med. Microbiol.* 16: 105-115; Moran *et al.* (1996b) *J. Endotoxin Res.* 3: 521-531).

Consequently, the identification of the genes involved in LPS synthesis and the study of their regulation is of considerable interest for a better understanding of the pathogenesis mechanisms used by these bacteria. Moreover, the use of gangliosides as therapeutic reagents, as well as the study of ganglioside function, would be facilitated by convenient and efficient methods of synthesizing desired gangliosides and ganglioside mimics. A combined enzymatic and chemical approach to synthesis of 3'-nLM$_1$ and 6'-nLM$_1$ has been described (Gaudino and Paulson (1994) *J. Am. Chem. Soc.* 116: 1149-1150). However, previously available enzymatic methods for ganglioside synthesis suffer from difficulties in efficiently producing enzymes in sufficient quantities, at a sufficiently low cost, for practical large-scale ganglioside synthesis. Thus, a need exists for new enzymes involved in ganglioside synthesis that are amenable to large-scale production. A need also exists for more efficient methods for synthesizing gangliosides. The present invention fulfills these and other needs.

Figures 1A-1C show lipooligosaccharide (*LOS*) outer core structures from *C. jejuni* O:19 strains. These structures were described by Aspinall *et al.* (1994) *Biochemistry* **33**, 241-249, and the portions showing similarity with the oligosaccharide portion of gangliosides are delimited by boxes. Figure 1A: *LOS* of *C. jejuni* O:19 serostrain (ATCC #43446) has structural similarity to the oligosaccharide portion of ganglioside GD1a. Figure 1B: *LOS* of *C. jejuni* O:19 strain OH4384 has structural similarity to the oligosaccharide portion of ganglioside GT1a. Figure 1C: *LOS* of *C. jejuni* OH4382 has structural similarity to the oligosaccharide portion of ganglioside GD3.

Figures 2A-2B show the genetic organization of the *cst-I* locus from OH4384 and comparison of the LOS biosynthesis loci from OH4384 and NCTC 11168. The distance between the scale marks is 1 kb. Figure 2A shows a schematic representation of the OH4384 *cst-I* locus, based on the nucleotide sequence which is available from GenBank (#AF130466). The partial *prfB* gene is somewhat similar to a peptide chain release factor (GenBank #AE000537) from *Helicobacter pylori,* while the *cysD* gene and the partial *cysN* gene are similar to *E. coli* genes encoding sulfate adenylyltransferase subunits (GenBank #AE000358). Figure 2B shows a schematic representation of the OH4384 LOS biosynthesis locus, which is based on the nucleotide sequence from GenBank (#AF130984). The nucleotide sequence of the OH4382 LOS biosynthesis locus is identical to that of OH4384 except for the *cgtA* gene, which is missing an "A" (see text and GenBank #AF167345). The sequence of the NCTC 11168 LOS biosynthesis locus is available from the Sanger Centre (URL:http//www.sanger.ac.uk/Projects/C_jejuni/). Corresponding homologous genes have the same number with a trailing "a" for the OH4384 genes and a trailing "b" for the NCTC 11168 genes. A gene unique to the OH4384 strain is shown in black and genes unique to NCTC 11168 are shown in grey. The OH4384 ORF's #5a and #10a are found as an in-frame fusion ORF (#5b/10b) in NCTC 11168 and are denoted with an asterisk (*). Proposed functions for each ORF are found in Table 4.

Figure 3 shows an alignment of the deduced amino acid sequences for the sialyltransferases. The OH4384 *cst-I* gene (first 300 residues), OH4384 *cst-II* gene (identical to OH4382 *cst-II*), O:19 (serostrain) *cst-II* gene (GenBank #AF167344) , NCTC 11168 *cst-II* gene and an *H. influenzae* putative ORF (GenBank #U32720) were aligned using the

4

ClustalX alignment program (Thompson *et al.* (1997) *Nucleic Acids Res.* 25, 4876-82). The shading was produced by the program GeneDoc (Nicholas, K. B., and Nicholas, H. B. (1997) URL: http://www.cris.com/~ketchup/genedoc.shtml).

5   Figure 4 shows a scheme for the enzymatic synthesis of ganglioside mimics using *C. jejuni* OH4384 glycosyltransferases. Starting from a synthetic acceptor molecule, a series of ganglioside mimics was synthesized with recombinant α-2,3-sialyltransferase (Cst-I), β-1,4-*N*-acetylgalactosaminyltransferase (CgtA), β-1,3-galactosyltransferase (CgtB), and a bi-functional α-2,3/α-2,8-sialyltransferase (Cst-II) using the sequences shown. All the products were analyzed by mass spectrometry and the observed monoisotopic masses

10   (shown in parentheses) were all within 0.02 % of the theoretical masses. The GM3, GD3, GM2 and GM1a mimics were also analyzed by NMR spectroscopy (see Table 4).

SUMMARY OF THE INVENTION

The present invention provides prokaryotic glycosyltransferase enzymes and nucleic acids that encode the enzymes. In one embodiment, the invention provides isolated

15   and/or recombinant nucleic acid molecules that include a polynucleotide sequence that encodes a polypeptide selected from the group consisting of:

a)   a polypeptide having lipid A biosynthesis acyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 350-1234 (ORF 2a) of

20   the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

b)   a polypeptide having glycosyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 1234-2487 (ORF 3a) of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

25   c)   a polypeptide having glycosyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 50 % identical to an amino acid sequence encoded by nucleotides 2786-3952 (ORF 4a) of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1 over a region at least about 100 amino acids in length;

d)   a polypeptide having β1,4-GalNAc transferase activity, wherein the GalNAc transferase polypeptide has an amino acid sequence that is at least about 77% identical to an amino acid sequence as set forth in SEQ ID NO:17 over a region at least about 50 amino acids in length;

e)   a polypeptide having β1,3-galactosyltransferase activity, wherein the galactosyltransferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:27 or SEQ ID NO:29 over a region at least about 50 amino acids in length;

f)   a polypeptide having either α2,3 sialyltransferase activity or both α2,3- and α2,8 sialyltransferase activity, wherein the polypeptide has an amino acid sequence that is at least about 66% identical over a region at least about 60 amino acids in length to an amino acid sequence as set forth in one or more of SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:10;

g)   a polypeptide having sialic acid synthase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 6924-7961 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

h)   a polypeptide having sialic acid biosynthesis activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 8021-9076 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

i)   a polypeptide having CMP-sialic acid synthetase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by nucleotides 9076-9738 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

j)   a polypeptide having acetyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by nucleotides 9729-10559 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1; and

6

k)   a polypeptide having glycosyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by a reverse complement of nucleotides 10557-11366 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1.

In presently preferred embodiments, the invention provides an isolated nucleic acid molecule that includes a polynucleotide sequence that encodes one or more polypeptides selected from the group consisting of: a) a sialyltransferase polypeptide that has both an α2,3 sialyltransferase activity and an α2,8 sialyltransferase activity, wherein the sialyltransferase polypeptide has an amino acid sequence that is at least about 76% identical to an amino acid sequence as set forth in SEQ ID NO:3 over a region at least about 60 amino acids in length; b) a GalNAc transferase polypeptide that has a β1,4-GalNAc transferase activity, wherein the GalNAc transferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:17 over a region at least about 50 amino acids in length; and c) a galactosyltransferase polypeptide that has β1,3-galactosyltransferase activity, wherein the galactosyltransferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:27 over a region at least about 50 amino acids in length.

Also provided by the invention are expression cassettes and expression vectors in which a glycosyltransferase nucleic acid of the invention is operably linked to a promoter and other control sequences that facilitate expression of the glycosyltransferases in a desired host cell. Recombinant host cells that express the glycosyltransferases of the invention are also provided.

The invention also provides isolated and/or recombinantly produced polypeptides selected from the group consisting of:

a)   a polypeptide having lipid A biosynthesis acyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 350-1234 (ORF 2a) of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

b)	a polypeptide having glycosyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 1234-2487 (ORF 3a) of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

c)	a polypeptide having glycosyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 50 % identical to an amino acid sequence encoded by nucleotides 2786-3952 (ORF 4a) of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1 over a region at least about 100 amino acids in length;

d)	a polypeptide having β1,4-GalNAc transferase activity, wherein the GalNAc transferase polypeptide has an amino acid sequence that is at least about 77% identical to an amino acid sequence as set forth in SEQ ID NO:17 over a region at least about 50 amino acids in length;

e)	a polypeptide having β1,3-galactosyltransferase activity, wherein the galactosyltransferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:27 or SEQ ID NO:29 over a region at least about 50 amino acids in length;

f)	a polypeptide having either α2,3 sialyltransferase activity or both α2,3 and α2,8 sialyltransferase activity, wherein the polypeptide has an amino acid sequence that is at least about 66% identical to an amino acid sequence as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or SEQ ID NO:10 over a region at least about 60 amino acids in length;

g)	a polypeptide having sialic acid synthase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 6924-7961 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

h)	a polypeptide having sialic acid biosynthesis activity, wherein the polypeptide comprises an amino acid sequence that is at least about 70% identical to an amino acid sequence encoded by nucleotides 8021-9076 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

8

i)   a polypeptide having CMP-sialic acid synthetase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by nucleotides 9076-9738 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1;

5   j)   a polypeptide having acetyltransferase activity, wherein the polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by nucleotides 9729-10559 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1; and

k)   a polypeptide having glycosyltransferase activity, wherein the
10   polypeptide comprises an amino acid sequence that is at least about 65% identical to an amino acid sequence encoded by a reverse complement of nucleotides 10557-11366 of the *LOS* biosynthesis locus of *C. jejuni* strain OH4384 as shown in SEQ ID NO:1.

In presently preferred embodiments, the invention provides
15   glycosyltransferase polypeptides including: a) a sialyltransferase polypeptide that has both an α2,3 sialyltransferase activity and an α2,8 sialyltransferase activity, wherein the sialyltransferase polypeptide has an amino acid sequence that is at least about 76% identical to an amino acid sequence as set forth in SEQ ID NO:3 over a region at least about 60 amino acids in length; b) a GalNAc transferase polypeptide that has a β1,4-GalNAc transferase
20   activity, wherein the GalNAc transferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:17 over a region at least about 50 amino acids in length; and c) a galactosyltransferase polypeptide that has β1,3-galactosyltransferase activity, wherein the galactosyltransferase polypeptide has an amino acid sequence that is at least about 75% identical to an amino acid sequence as
25   set forth in SEQ ID NO:27 or SEQ ID NO:29 over a region at least about 50 amino acids in length.

The invention also provides reaction mixtures for the synthesis of a sialylated oligosaccharide. The reaction mixtures include a sialyltransferase polypeptide which has both an α2,3 sialyltransferase activity and an α2,8 sialyltransferase activity. Also present in
30   the reaction mixtures are a galactosylated acceptor moiety and a sialyl-nucleotide sugar. The

sialyltransferase transfers a first sialic acid residue from the sialyl-nucleotide sugar (*e.g.*, CMP-sialic acid) to the galactosylated acceptor moiety in an α2,3 linkage, and further adds a second sialic acid residue to the first sialic acid residue in an α2,8 linkage.

In another embodiment, the invention provides methods for synthesizing a sialylated oligosaccharide. These methods involve incubating a reaction mixture that includes a sialyltransferase polypeptide which has both an α2,3 sialyltransferase activity and an α2,8 sialyltransferase activity, a galactosylated acceptor moiety, and a sialyl-nucleotide sugar, under suitable conditions wherein the sialyltransferase polypeptide transfers a first sialic acid residue from the sialyl-nucleotide sugar to the galactosylated acceptor moiety in an α2,3 linkage, and further transfers a second sialic acid residue to the first sialic acid residue in an α2,8 linkage.

DETAILED DESCRIPTION

Definitions

The glycosyltransferases, reaction mixtures, and methods of the invention are useful for transferring a monosaccharide from a donor substrate to an acceptor molecule. The addition generally takes place at the non-reducing end of an oligosaccharide or carbohydrate moiety on a biomolecule. Biomolecules as defined here include, but are not limited to, biologically significant molecules such as carbohydrates, proteins (*e.g.*, glycoproteins), and lipids (*e.g.*, glycolipids, phospholipids, sphingolipids and gangliosides).

The following abbreviations are used herein:

Ara	= arabinosyl;
Fru	= fructosyl;
Fuc	= fucosyl;
Gal	= galactosyl;
GalNAc	= N-acetylgalactosaminyl;
Glc	= glucosyl;
GlcNAc	= N-acetylglucosaminyl;
Man	= mannosyl; and
NeuAc	= sialyl (N-acetylneuraminyl).

The term "sialic acid" refers to any member of a family of nine-carbon carboxylated sugars. The most common member of the sialic acid family is N-acetyl-neuraminic acid (2-keto-5-acetamindo-3,5-dideoxy-D-glycero-D-galactononulopyranos-1-onic acid (often abbreviated as Neu5Ac, NeuAc, or NANA). A second member of the

5   family is N-glycolyl-neuraminic acid (Neu5Gc or NeuGc), in which the N-acetyl group of NeuAc is hydroxylated. A third sialic acid family member is 2-keto-3-deoxy-nonulosonic acid (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* 261: 11550-11557; Kanamori *et al.* (1990) *J. Biol. Chem.* 265: 21811-21819. Also included are 9-substituted sialic acids such as a 9-O-C_1-C_6 acyl-Neu5Ac like 9-O-lactyl-Neu5Ac or 9-O-acetyl-Neu5Ac, 9-deoxy-9-fluoro-

10   Neu5Ac and 9-azido-9-deoxy-Neu5Ac. For review of the sialic acid family, *see, e.g.*, Varki (1992) *Glycobiology* 2: 25-40; *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, New York (1992); Schauer, *Methods in Enzymology*, 50: 64-89 (1987), and Schaur, *Advances in Carbohydrate Chemistry and Biochemistry*, 40: 131-234.The synthesis and use of sialic acid compounds in a sialylation procedure is disclosed in

15   international application WO 92/16640, published October 1, 1992.

    Donor substrates for glycosyltransferases are activated nucleotide sugars. Such activated sugars generally consist of uridine and guanosine diphosphates, and cytidine monophosphate derivatives of the sugars in which the nucleoside diphosphate or monophosphate serves as a leaving group. Bacterial, plant, and fungal systems can

20   sometimes use other activated nucleotide sugars.

    Oligosaccharides are considered to have a reducing end and a non-reducing end, whether or not the saccharide at the reducing end is in fact a reducing sugar. In accordance with accepted nomenclature, oligosaccharides are depicted herein with the non-reducing end on the left and the reducing end on the right.

25   All oligosaccharides described herein are described with the name or abbreviation for the non-reducing saccharide (*e.g.*, Gal), followed by the configuration of the glycosidic bond (α or β), the ring bond, the ring position of the reducing saccharide involved in the bond, and then the name or abbreviation of the reducing saccharide (*e.g.*, GlcNAc). The linkage between two sugars may be expressed, for example, as 2,3, 2→3, or (2,3). Each

30   saccharide is a pyranose or furanose.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence affects transcription and/or translation of the nucleic acid corresponding to the second sequence.

A "heterologous polynucleotide" or a "heterologous nucleic acid", as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous glycosyltransferase gene in a host cell includes a glycosyltransferase gene that is endogenous to the particular host cell but has been modified. Modification of the heterologous sequence may occur, e.g., by treating the DNA with a restriction enzyme to generate a DNA fragment that is capable of being operably linked to a promoter. Techniques such as site-directed mutagenesis are also useful for modifying a heterologous sequence.

The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells also include those that contain genes that are found in the native form of the cell, but are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques known to those of skill in the art.

A "recombinant nucleic acid" is a nucleic acid that is in a form that is altered from its natural state. For example, the term "recombinant nucleic acid" includes a coding region that is operably linked to a promoter and/or other expression control region,

12

processing signal, another coding region, and the like., to which the nucleic acid is not linked in its naturally occurring form. A "recombinant nucleic acid" also includes, for example, a coding region or other nucleic acid in which one or more nucleotides have been substituted, deleted, inserted, compared to the corresponding naturally occurring nucleic acid. The

5   modifications include those introduced by *in vitro* manipulation, *in vivo* modification, synthesis methods, and the like.

      A "recombinantly produced polypeptide" is a polypeptide that is encoded by a recombinant and/or heterologous nucleic acid. For example, a polypeptide that is expressed from a *C. jejuni* glycosyltransferase-encoding nucleic acid which is introduced into *E. coli* is

10   a "recombinantly produced polypeptide." A protein expressed from a nucleic acid that is operably linked to a non-native promoter is one example of a "recombinantly produced polypeptide. Recombinantly produced polypeptides of the invention can be used to synthesize gangliosides and other oligosaccharides in their unpurified form (*e.g.*, as a cell lysate or an intact cell), or after being completely or partially purified.

15   A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with nucleic acid elements that are capable of affecting expression of a structural gene in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes a nucleic acid to

20   be transcribed (*e.g.*, a nucleic acid encoding a desired polypeptide), and a promoter. Additional factors necessary or helpful in effecting expression may also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence

25   gene expression, can also be included in an expression cassette.

      A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (*e.g.*, polypeptide) respectively.

      The term "isolated" is meant to refer to material that is substantially or

30   essentially free from components which normally accompany the material as found in its

native state. Typically, isolated proteins or nucleic acids of the invention are at least about 80% pure, usually at least about 90%, and preferably at least about 95% pure. Purity or homogeneity can be indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a protein or nucleic acid sample, followed by

5   visualization upon staining. For certain purposes high resolution will be needed and HPLC or a similar means for purification utilized. An "isolated" enzyme, for example, is one which is substantially or essentially free from components which interfere with the activity of the enzyme. An "isolated nucleic acid" includes, for example, one that is not present in the chromosome of the cell in which the nucleic acid naturally occurs.

10   The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

15   The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least 60%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity

20   exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

For sequence comparison, typically one sequence acts as a reference

25   sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

14

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally*, *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)).

Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschuel *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). For example, the comparisons can be performed using a BLASTN Version 2.0 algorithm with a wordlength (W) of 11, G=5, E=2, q= -2, and r = 1., and a comparison of both strands. For amino acid sequences, the BLASTP Version 2.0 algorithm can be used, with the default values of wordlength (W) of 3, G=11, E=1, and a BLOSUM62 substitution matrix. (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

The phrase "hybridizing specifically to", refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at Tm, 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

A further indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

The phrases "specifically binds to a protein" or "specifically immunoreactive with", when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind preferentially to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires

16

an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

"Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. One of skill will appreciate that many conservative variations of the fusion proteins and nucleic acid which encode the fusion proteins yield essentially identical products. For example, due to the degeneracy of the

17

genetic code, "silent substitutions" (*i.e.,* substitutions of a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. As described herein, sequences are preferably optimized for expression in a particular host cell used to produce the enzymes (*e.g.,* yeast,

5   human, and the like). Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties (*see,* the definitions section, *supra*), are also readily identified as being highly similar to a particular amino acid sequence, or to a particular nucleic acid sequence which encodes an amino acid. Such conservatively substituted variations of any particular

10   sequence are a feature of the present invention. *See also,* Creighton (1984) *Proteins,* W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations".

Description of the Preferred Embodiments

15   The present invention provides novel glycosyltransferase enzymes, as well as other enzymes that are involved in enzyme-catalyzed oligosaccharide synthesis. The glycosyltransferases of the invention include sialyltransferases, including a bifunctional sialyltransferase that has both an $\alpha2,3$ and an $\alpha2,8$ sialyltransferase activity. Also provided are $\beta1,3$-galactosyltransferases, $\beta1,4$-GalNAc transferases, sialic acid synthases, CMP-sialic

20   acid synthetases, acetyltransferases, and other glycosyltransferases. The enzymes of the invention are prokaryotic enzymes, include those involved in the biosynthesis of lipooligosaccharides (LOS) in various strains of *Campylobacter jejuni*. The invention also provides nucleic acids that encode these enzymes, as well as expression cassettes and expression vectors for use in expressing the glycosyltransferases. In additional embodiments,

25   the invention provides reaction mixtures and methods in which one or more of the enzymes is used to synthesize an oligosaccharide.

                The glycosyltransferases of the invention are useful for several purposes. For example, the glycosyltransferases are useful as tools for the chemo-enzymatic syntheses of oligosaccharides, including gangliosides and other oligosaccharides that have biological

30   activity. The glycosyltransferases of the invention, and nucleic acids that encode the

18

glycosyltransferases, are also useful for studies of the pathogenesis mechanisms of organisms that synthesize ganglioside mimics, such as *C. jejuni*. The nucleic acids can be used as probes, for example, to study expression of the genes involved in ganglioside mimetic synthesis. Antibodies raised against the glycosyltransferases are also useful for

5   analyzing the expression patterns of these genes that are involved in pathogenesis. The nucleic acids are also useful for designing antisense oligonucleotides for inhibiting expression of the *Campylobacter* enzymes that are involved in the biosynthesis of ganglioside mimics that can mask the pathogens from the host's immune system.

     The glycosyltransferases of the invention provide several advantages over

10   previously available glycosyltransferases. Bacterial glycosyltransferases such as those of the invention can catalyze the formation of oligosaccharides that are identical to the corresponding mammalian structures. Moreover, bacterial enzymes are easier and less expensive to produce in quantity, compared to mammalian glycosyltransferases. Therefore, bacterial glycosyltransferases such as those of the present invention are attractive

15   replacements for mammalian glycosyltransferases, which can be difficult to obtain in large amounts. That the glycosyltransferases of the invention are of bacterial origin facilitates expression of large quantities of the enzymes using relatively inexpensive prokaryotic expression systems. Typically, prokaryotic systems for expression of polypeptide products involves a much lower cost than expression of the polypeptides in mammalian cell culture

20   systems.

     Moreover, the novel bifunctional sialyltransferases of the invention simplify the enzymatic synthesis of biologically important molecules, such as gangliosides, that have a sialic acid attached by an $\alpha2,8$ linkage to a second sialic acid, which in turn is $\alpha2,3$-linked to a galactosylated acceptor. While previous methods for synthesizing these structures

25   required two separate sialyltransferases, only one sialyltransferase is required when the bifunctional sialyltransferase of the present invention is used. This avoids the costs associated with obtaining a second enzyme, and can also reduce the number of steps involved in synthesizing these compounds.

A.   *Glycosyltransferases and associated enzymes*

The present invention provides prokaryotic glycosyltransferase polypeptides, as well as other enzymes that are involved in the glycosyltransferase-catalyzed synthesis of oligosaccharides, including gangliosides and ganglioside mimics. In presently preferred embodiments, the polypeptides include those that are encoded by open reading frames within the lipooligosaccharide (*LOS*) locus of *Campylobacter* species (Figure 1). Included among the enzymes of the invention are glycosyltransferases, such as sialyltransferases (including a bifunctional sialyltransferase), β1,4-GalNAc transferases, and β1,3-galactosyltransferases, among other enzymes as described herein. Also provided are accessory enzymes such as, for example, CMP-sialic acid synthetase, sialic acid synthase, acetyltransferase, an acyltransferase that is involved in lipid A biosynthesis, and an enzyme involved in sialic acid biosynthesis.

The glycosyltransferases and accessory polypeptides of the invention can be purified from natural sources, *e.g.*, prokaryotes such as *Campylobacter* species. In presently preferred embodiments, the glycosyltransferases are obtained from *C. jejuni*, in particular from *C. jejuni* serotype O:19, including the strains OH4384 and OH4382. Also provided are glycosyltransferases and accessory enzymes obtained from *C. jejuni* serotypes O:10, O:41, and O:2. Methods by which the glycosyltransferase polypeptides can be purified include standard protein purification methods including, for example, ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see, generally*, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982) Deutscher, *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*., Academic Press, Inc. N.Y. (1990)).

In presently preferred embodiments, the glycosyltransferase and accessory enzyme polypeptides of the invention are obtained by recombinant expression using the glycosyltransferase- and accessory enzyme-encoding nucleic acids described herein. Expression vectors and methods for producing the glycosyltransferases are described in detail below.

In some embodiments, the glycosyltransferase polypeptides are isolated from their natural milieu, whether recombinantly produced or purified from their natural cells. Substantially pure compositions of at least about 90 to 95% homogeneity are preferred for

some applications, and 98 to 99% or more homogeneity are most preferred. Once purified, partially or to homogeneity as desired, the polypeptides may then be used (*e.g.*, as immunogens for antibody production or for synthesis of oligosaccharides, or other uses as described herein or apparent to those of skill in the art). The glycosyltransferases need not,

5 however, be even partially purified for use to synthesize a desired saccharide structure. For example, the invention provides recombinantly produced enzymes that are expressed in a heterologous host cell and/or from a recombinant nucleic acid. Such enzymes of the invention can be used when present in a cell lysate or an intact cell, as well as in purified form.

10   *i. Sialyltransferases*

        In some embodiments, the invention provides sialyltransferase polypeptides. The sialyltransferases have an $\alpha 2,3$-sialyltransferase activity, and in some cases also have an $\alpha 2,8$ sialyltransferase activity. These bifunctional sialyltransferases, when placed in a reaction mixture with a suitable saccharide acceptor (*e.g.*, a saccharide having a terminal

15 galactose) and a sialic acid donor (*e.g.*, CMP-sialic acid) can catalyze the transfer of a first sialic acid from the donor to the acceptor in an $\alpha 2,3$ linkage. The sialyltransferase then catalyzes the transfer of a second sialic acid from a sialic acid donor to the first sialic acid residue in an $\alpha 2,8$ linkage. This type of Sia$\alpha 2,8$-Sia$\alpha 2,3$-Gal structure is often found in gangliosides, including GD3 and GT1a as shown in Figure 4.

20   Examples of bifunctional sialyltransferases of the invention are those that are found in *Campylobacter* species, such as *C. jejuni*. A presently preferred bifunctional sialyltransferase of the invention is that of the *C. jejuni* serotype O:19. One example of a bifunctional sialyltransferase is that of *C. jejuni* strain OH 4384; this sialyltransferase has an amino acid sequence as shown in SEQ ID NO:3. Other bifunctional sialyltransferases of the

25 invention generally have an amino acid sequence that is at least about 76% identical to the amino acid sequence of the *C. jejuni* OH4384 bifunctional sialyltransferase over a region at least about 60 amino acids in length. More preferably, the sialyltransferases of the invention are at least about 85% identical to the OH 4384 sialyltransferase amino acid sequence, and still more preferably at least about 95% identical to the amino acid sequence of SEQ ID

30 NO:3, over a region of at least 60 amino acids in length. In presently preferred embodiments,

the region of percent identity extends over a region longer than 60 amino acids. For example, in more preferred embodiments, the region of similarity extends over a region of at least about 100 amino acids in length, more preferably a region of at least about 150 amino acids in length, and most preferably over the full length of the sialyltransferase. Accordingly, the

5   bifunctional sialyltransferases of the invention include polypeptides that have either or both the α2,3- and α2,8-sialyltransferase activity and are at least about 65% identical, more preferably at least about 70% identical, more preferably at least about 80% identical, and most preferably at least about 90% identical to the amino acid sequence of the *C. jejuni* OH 4384 CstII sialyltransferase (SEQ ID NO:3) over a region of the polypeptide that is required

10   to retain the respective sialyltransferase activities. In some embodiments, the bifunctional sialyltransferases of the invention are identical to *C. jejuni* OH 4384 CstII sialyltransferase over the entire length of the sialyltransferase.

The invention also provides sialyltransferases that have α2,3 sialyltransferase activity, but little or no α2,8 sialyltransferase activity. For example, CstII sialyltransferase of

15   the *C. jejuni* O:19 serostrain (SEQ ID NO:9) differs from that of strain OH 4384 by eight amino acids, but nevertheless substantially lacks α2,8 sialyltransferase activity (Figure 3). The corresponding sialyltransferase from the O:2 serotype strain NCTC 11168 (SEQ ID NO:10) is 52% identical to that of OH4384, and also has little or no α2,8-sialyltranfserase activity. Sialyltransferases that are substantially identical to the CstII sialyltransferase of *C.*

20   *jejuni* strain O:10 (SEQ ID NO:5) and O:41 (SEQ ID NO:7) are also provided. The sialyltransferases of the invention include those that are at least about 65% identical, more preferably at least about 70% identical, more preferably at least about 80% identical, and most preferably at least about 90% identical to the amino acid sequences of the *C. jejuni* O:10 (SEQ ID NO:5), O:41 (SEQ ID NO:7), O:19 serostrain (SEQ ID NO:9), or O:2

25   serotype strain NCTC 11168 (SEQ ID NO:10). The sialyltransferases of the invention, in some embodiments, have an amino acid sequence that is identical to that of the O:10, O:41, O:19 serostrain or NCTC 11168 *C. jejuni* strains.

The percent identities can be determined by inspection, for example, or can be determined using an alignment algorithm such as the BLASTP Version 2.0 algorithm using

the default parameters, such as a wordlength (W) of 3, G=11, E=1, and a BLOSUM62 substitution matrix.

Sialyltransferases of the invention can be identified, not only by sequence comparison, but also by preparing antibodies against the *C. jejuni* OH4384 bifunctional

5   sialyltransferase, or other sialyltransferases provided herein, and determining whether the antibodies are specifically immunoreactive with a sialyltransferase of interest. To obtain a bifunctional sialyltransferase in particular, one can identify an organism that is likely to produce a bifunctional sialyltransferase by determining whether the organism displays both α2,3 and α2,8-sialic acid linkages on its cell surfaces. Alternatively, or in addition, one can

10   simply do enzyme assays of an isolated sialyltransferase to determine whether both sialyltransferase activities are present.

2.   *β1,4-GalNAc transferase*

The invention also provides β1,4-GalNAc transferase polypeptides (*e.g.,* *CgtA*). The β1,4-GalNAc transferases of the invention, when placed in a reaction mixture,

15   catalyze the transfer of a GalNAc residue from a donor (*e.g.,* UDP-GalNAc) to a suitable acceptor saccharide (typically a saccharide that has a terminal galactose residue). The resulting structure, GalNAcβ1,4-Gal-, is often found in gangliosides and other sphingoids, among many other saccharide compounds. For example, the CgtA transferase can catalyze the conversion of the ganglioside GM3 to GM2 (Figure 4).

20   Examples of the β1,4-GalNAc transferases of the invention are those that are produced by *Campylobacter* species, such as *C. jejuni.* One example of a β1,4-GalNAc transferase polypeptide is that of *C. jejuni* strain OH4384, which has an amino acid sequence as shown in SEQ ID NO:17. The β1,4-GalNAc transferases of the invention generally include an amino acid sequence that is at least about 75% identical to an amino acid

25   sequence as set forth in SEQ ID NO:17 over a region at least about 50 amino acids in length. More preferably, the β1,4-GalNAc transferases of the invention are at least about 85% identical to this amino acid sequence, and still more preferably are at least about 95% identical to the amino acid sequence of SEQ ID NO:17, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends

30   over a longer region than 50 amino acids, more preferably over a region of at least about 100

23

amino acids, and most preferably over the full length of the GalNAc transferase. Accordingly, the β1,4-GalNAc transferases of the invention include polypeptides that have β1,4-GalNAc transferase activity and are at least about 65% identical, more preferably at least about 70% identical, more preferably at least about 80% identical, and most preferably

5 at least about 90% identical to the amino acid sequence of the *C. jejuni* OH 4384 β1,4-GalNAc transferases (SEQ ID NO:17) over a region of the polypeptide that is required to retain the β1,4-GalNAc transferase activity. In some embodiments, the β1,4-GalNAc transferases of the invention are identical to *C. jejuni* OH 4384 β1,4-GalNAc transferase over the entire length of the β1,4-GalNAc transferase.

10   . .- Again, the percent identities can be tletermined by inspection, for example, or can be determined using an alignment algorithm such as the BLASTP Version 2.0 algorithm with a wordlength (W) of 3, G=11, E=1, and a BLOSUM62 substitution matrix.

One can also identify β1,4-GalNAc transferases of the invention by immunoreactivity. For example, one can prepare antibodies against the *C. jejuni* OH4384

15 β1,4-GalNAc transferase of SEQ ID NO:17 and determine whether the antibodies are specifically immunoreactive with a β1,4-GalNAc transferase of interest.

3.   *β1,3-Galactosyltransferases*

Also provided by the invention are β1,3-galactosyltransferases (CgtB). When placed in a suitable reaction medium, the β1,3-galactosyltransferases of the invention

20 catalyze the transfer of a galactose residue from a donor (*e.g.*, UDP-Gal) to a suitable saccharide acceptor (*e.g.*, saccharides having a terminal GalNAc residue). Among the reactions catalyzed by the β1,3-galactosyltransferases is the transfer of a galactose residue to the oligosaccharide moiety of GM2 to form the GM1a oligosaccharide moiety.

Examples of the β1,3-galactosyltransferases of the invention are those

25 produced by *Campylobacter* species, such as *C. jejuni*. For example, one β1,3-galactosyl-transferase of the invention is that of *C. jejuni* strain OH4384, which has the amino acid sequence shown in SEQ ID NO:27.

Another example of a β1,3-galactosyltransferase of the invention is that of the *C. jejuni* O:2 serotype strain NCTC 11168. The amino acid sequence of this

30 galactosyltransferase is set forth in SEQ ID NO:29. This galactosyltransferase expresses well

24

in *E. coli*, for example, and exhibits a high amount of soluble activity. Moreover, unlike the OH4384 CgtB, which can add more than one galactose if a reaction mixture contains an excess of donor and is incubated for a sufficiently long period of time, the NCTC 11168 β1,3-galactose does not have a significant amount of polygalactosyltransferase activity. For some applications, the polygalactosyltransferase activity of the OH4384 enzyme is desirable, but in other applications such as synthesis of GM1 mimics, addition of only one terminal galactose is desirable.

The β1,3-galactosyltransferases of the invention generally have an amino acid sequence that is at least about 75% identical to an amino acid sequence of the OH 4384 or NCTC 11168 CgtB as set forth in SEQ ID NO:27 and SEQ ID NO:29, respectively, over a region at least about 50 amino acids in length. More preferably, the β1,3-galactosyltransferases of the invention are at least about 85% identical to either of these amino acid sequences, and still more preferably are at least about 95% identical to the amino acid sequences of SEQ ID NO:27 or SEQ ID NO:29, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 50 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the galactosyltransferase. Accordingly, the β1,3-galactosyltransferases of the invention include polypeptides that have β1,3-galactosyltransferase activity and are at least about 65% identical, more preferably at least about 70% identical, more preferably at least about 80% identical, and most preferably at least about 90% identical to the amino acid sequence of the *C. jejuni* OH4384 β1,3-galactosyltransferase (SEQ ID NO:27) or the NCTC 11168 galactosyltransferase (SEQ ID NO:29) over a region of the polypeptide that is required to retain the β1,3-galactosyltransferase activity. In some embodiments, the β1,3-galactosyltransferase of the invention are identical to *C. jejuni* OH 4384 or NCTC 11168 β1,3-galactosyltransferase over the entire length of the β1,3-galactosyltransferase.

The percent identities can be determined by inspection, for example, or can be determined using an alignment algorithm such as the BLASTP Version 2.0 algorithm with a wordlength (W) of 3, G=11, E=1, and a BLOSUM62 substitution matrix.

25

The β1,3-galactosyltransferases of the invention can be obtained from the respective *Campylobacter* species, or can be produced recombinantly. One can identify the glycosyltransferases by assays of enzymatic activity, for example, or by detecting specific immunoreactivity with antibodies raised against the *C. jejuni* OH4384 β1,3-galactosyltransferase having an amino acid sequence as set forth in SEQ ID NO:27 or the *C. jejuni* NCTC 11168 β1,3 galactosyltransferase as set forth in SEQ ID NO:29.

4. Additional enzymes involved in LOS biosynthetic pathway

The present invention also provides additional enzymes that are involved in the biosynthesis of oligosaccharides such as those found on bacterial lipooligosaccharides. For example, enzymes involved in the synthesis of CMP-sialic acid, the donor for sialyltransferases, are provided. A sialic acid synthase is encoded by open reading frame (ORF) 8a of *C. jejuni* strain OH 4384 (SEQ ID NO:35) and by open reading frame 8b of strain NCTC 11168 (*see*, Table 3). Another enzyme involved in sialic acid synthesis is encoded by ORF 9a of OH 4384 (SEQ ID NO:36) and 9b of NCTC 11168. A CMP-sialic acid synthetase is encoded by ORF 10a (SEQ ID NO:37) and 10b of OH 4384 and NCTC 11168, respectively.

The invention also provides an acyltransferase that is involved in lipid A biosynthesis. This enzyme is encoded by open reading frame 2a of *C. jejuni* strain OH4384 (SEQ ID NO:32) and by open reading frame 2B of strain NCTC 11168. An acetyltransferase is also provided; this enzyme is encoded by ORF 11a of strain OH 4384 (SEQ ID NO:38); no homolog is found in the *LOS* biosynthesis locus of strain NCTC 11168.

Also provided are three additional glycosyltransferases. These enzymes are encoded by ORFs 3a (SEQ ID NO:33), 4a (SEQ ID NO:34), and 12a (SEQ ID NO:39) of strain OH 4384 and ORFs 3b, 4b, and 12b of strain NCTC 11168.

The invention includes, for each of these enzymes, polypeptides that include an an amino acid sequence that is at least about 75% identical to an amino acid sequence as set forth herein over a region at least about 50 amino acids in length. More preferably, the enzymes of the invention are at least about 85% identical to the respective amino acid sequence, and still more preferably are at least about 95% identical to the amino acid sequence, over a region of at least 50 amino acids in length. In presently preferred

26

embodiments, the region of percent identity extends over a longer region than 50 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the enzyme. Accordingly, the enzymes of the invention include polypeptides that have the respective activity and are at least about 65% identical, more

5   preferably at least about 70% identical, more preferably at least about 80% identical, and most preferably at least about 90% identical to the amino acid sequence of the corresponding enzyme as set forth herein over a region of the polypeptide that is required to retain the respective enzymatic activity. In some embodiments, the enzymes of the invention are identical to the corresponding *C. jejuni* OH 4384 enzymes over the entire length of the

10   enzyme.

B.   *Nucleic acids that encode glycosyltransferases and related enzymes*

The present invention also provides isolated and/or recombinant nucleic acids that encode the glycosyltransferases and other enzymes of the invention. The glycosyltransferase-encoding nucleic acids of the invention are useful for several purposes,

15   including the recombinant expression of the corresponding glycosyltransferase polypeptides, and as probes to identify nucleic acids that encode other glycosyltransferases and to study regulation and expression of the enzymes.

Nucleic acids of the invention include those that encode an entire glycosyltransferase enzyme such as those described above, as well as those that encode a

20   subsequence of a glycosyltransferase polypeptide. For example, the invention includes nucleic acids that encode a polypeptide which is not a full-length glycosyltransferase enzyme, but nonetheless has glycosyltransferase activity. The nucleotide sequences of the *LOS* locus of *C. jejuni* strain OH4384 is provided herein as SEQ ID NO:1, and the respective reading frames are identified. Additional nucleotide sequences are also provided, as

25   discussed below. The invention includes not only nucleic acids that include the nucleotide sequences as set forth herein, but also nucleic acids that are substantially identical to, or substantially complementary to, the exemplified embodiments. For example, the invention includes nucleic acids that include a nucleotide sequence that is at least about 70% identical to one that is set forth herein, more preferably at least 75%, still more preferably at least

30   80%, more preferably at least 85%, still more preferably at least 90%, and even more

27

preferably at least about 95% identical to an exemplified nucleotide sequence. The region of identity extends over at least about 50 nucleotides, more preferably over at least about 100 nucleotides, still more preferably over at least about 500 nucleotides. The region of a specified percent identity, in some embodiments, encompasses the coding region of a

5   sufficient portion of the encoded enzyme to retain the respective enzyme activity. The specified percent identity, in preferred embodiments, extends over the full length of the coding region of the enzyme.

The nucleic acids that encode the glycosyltransferases of the invention can be obtained using methods that are known to those of skill in the art. Suitable nucleic acids

10   (*e.g.,* cDNA, genomic, or subsequences (probes)) can be cloned, or amplified by *in vitro* methods such as the polymerase chain reaction (PCR), the ligase chain reaction (LCR), the transcription-based amplification system (TAS), the self-sustained sequence replication system (SSR). A wide variety of cloning and *in vitro* amplification methodologies are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct

15   persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook *et al.*); *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds.,

20   Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1994 Supplement) (Ausubel); Cashion *et al.*, U.S. patent number 5,017,478; and Carr, European Patent No. 0,246,864. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, (1987) U.S. Patent No. 4,683,202; *PCR Protocols A*

25   *Guide to Methods and Applications* (Innis *et al.*, eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (October 1, 1990) *C&EN* 36-47; *The Journal Of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem.,* 35: 1826; Landegren *et al.,* (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8:

30   291-294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer *et al.* (1990) *Gene* 89: 117.

Improved methods of cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039.

Nucleic acids that encode the glycosyltransferase polypeptides of the invention, or subsequences of these nucleic acids, can be prepared by any suitable method as

5   described above, including, for example, cloning and restriction of appropriate sequences. As an example, one can obtain a nucleic acid that encodes a glycosyltransferase of the invention by routine cloning methods. A known nucleotide sequence of a gene that encodes the glycosyltransferase of interest, such as are described herein, can be used to provide probes that specifically hybridize to a gene that encodes a suitable enzyme in a genomic

10   DNA sample, or to a mRNA in a total RNA sample (*e.g.*, in a Southern or Northern blot). Preferably, the samples are obtained from prokaryotic organisms, such as *Campylobacter* species. Examples of *Campylobacter species* of particular interest include *C. jejuni*. Many *C. jejuni* O:19 strains synthesize ganglioside mimics and are useful as a source of the glycosyltransferases of the invention.

15   Once the target glycosyltransferase nucleic acid is identified, it can be isolated according to standard methods known to those of skill in the art (*see, e.g.,* Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual, 2nd Ed., Vols. 1-3*, Cold Spring Harbor Laboratory; Berger and Kimmel (1987) *Methods in Enzymology, Vol. 152: Guide to Molecular Cloning Techniques,* San Diego: Academic Press, Inc.; or Ausubel *et al.* (1987)

20   *Current Protocols in Molecular Biology,* Greene Publishing and Wiley-Interscience, New York).

A nucleic acid that encodes a glycosyltransferase of the invention can also be cloned by detecting its expressed product by means of assays based on the physical, chemical, or immunological properties. For example, one can identify a cloned bifunctional

25   sialyltransferase-encoding nucleic acid by the ability of a polypeptide encoded by the nucleic acid to catalyze the coupling of a sialic acid in an α2,3-linkage to a galactosylated acceptor, followed by the coupling of a second sialic acid residue to the first sialic acid in an α2,8 linkage. Similarly, one can identify a cloned nucleic acid that encodes a β1,4-GalNAc transferase or a β1,3-galactosyltransferase by the ability of the encoded polypeptide to

30   catalyze the transfer of a GalNAc residue from UDP-GalNAc, or a galactose residue from

UDP-Gal, respectively, to a suitable acceptor. Suitable assay conditions are known in the art, and include those that are described in the Examples. Other physical properties of a polypeptide expressed from a particular nucleic acid can be compared to properties of known glycosyltransferase polypeptides of the invention, such as those described herein, to provide

5   another method of identifying nucleic acids that encode glycosyltransferases of the invention. Alternatively, a putative glycosyltransferase gene can be mutated, and its role as a glycosyltransferase established by detecting a variation in the ability to produce the respective glycoconjugate.

In other embodiments, glycosyltransferase-encoding nucleic acids can be

10   cloned using DNA amplification methods such as polymerase chain reaction (PCR). Thus, for example, the nucleic acid sequence or subsequence is PCR amplified, preferably using a sense primer containing one restriction site (*e.g., Xba*I) and an antisense primer containing another restriction site (*e.g., Hind*III). This will produce a nucleic acid encoding the desired glycosyltransferase amino acid sequence or subsequence and having terminal restriction

15   sites. This nucleic acid can then be easily ligated into a vector containing a nucleic acid encoding the second molecule and having the appropriate corresponding restriction sites. Suitable PCR primers can be determined by one of skill in the art using the sequence information provided herein. Appropriate restriction sites can also be added to the nucleic acid encoding the glycosyltransferase of the invention, or amino acid subsequence, by site-

20   directed mutagenesis. The plasmid containing the glycosyltransferase-encoding nucleotide sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods.

Examples of suitable primers suitable for amplification of the

25   glycosyltransferase-encoding nucleic acids of the invention are shown in Table 2; some of the primer pairs are designed to provide a 5' *Nde*I restriction site and a 3' *Sal*I site on the amplified fragment. The plasmid containing the enzyme-encoding sequence or subsequence is cleaved with the appropriate restriction endonuclease and then ligated into an appropriate vector for amplification and/or expression according to standard methods.

As an alternative to cloning a glycosyltransferase-encoding nucleic acid, a suitable nucleic acid can be chemically synthesized from a known sequence that encodes a glycosyltransferase of the invention. Direct chemical synthesis methods include, for example, the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68: 109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; and the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill would recognize that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences. Alternatively, subsequences may be cloned and the appropriate subsequences cleaved using appropriate restriction enzymes. The fragments can then be ligated to produce the desired DNA sequence.

In some embodiments, it may be desirable to modify the enzyme-encoding nucleic acids. One of skill will recognize many ways of generating alterations in a given nucleic acid construct. Such well-known methods include site-directed mutagenesis, PCR amplification using degenerate oligonucleotides, exposure of cells containing the nucleic acid to mutagenic agents or radiation, chemical synthesis of a desired oligonucleotide (*e.g.*, in conjunction with ligation and/or cloning to generate large nucleic acids) and other well-known techniques. *See, e.g.*, Giliman and Smith (1979) *Gene* 8:81-97, Roberts *et al.* (1987) *Nature* 328: 731-734.

In a presently preferred embodiment, the recombinant nucleic acids present in the cells of the invention are modified to provide preferred codons which enhance translation of the nucleic acid in a selected organism (*e.g., E. coli* preferred codons are substituted into a coding nucleic acid for expression in *E. coli*).

The present invention includes nucleic acids that are isolated (*i.e.*, not in their native chromosomal location) and/or recombinant (*i.e.*, modified from their original form, present in a non-native organism, etc.).

1. Sialyltransferases

The invention provides nucleic acids that encode sialyltransferases such as those described above. In some embodiments, the nucleic acids of the invention encode bifunctional sialyltransferase polypeptides that have both an α2,3 sialyltransferase activity and an α2,8 sialyltransferase activity. These sialyltransferase nucleic acids encode a sialyltransferase polypeptide that has an amino acid sequence that is at least about 76% identical to an amino acid sequence as set forth in SEQ ID NO:3 over a region at least about 60 amino acids in length. More preferably the sialyltransferases encoded by the nucleic acids of the invention are at least about 85% identical to the amino acid sequence of SEQ ID NO:3, and still more preferably at least about 95% identical to the amino acid sequence of SEQ ID NO:3, over a region of at least 60 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 60 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the sialyltransferase. In a presently preferred embodiment, the sialyltransferase-encoding nucleic acids of the invention encode a polypeptide having the amino acid sequence as shown in SEQ ID NO:3.

An example of a nucleic acid of the invention is an isolated and/or recombinant form of a bifunctional sialyltransferase-encoding nucleic acid of *C. jejuni* OH4384. The nucleotide sequence of this nucleic acid is shown in SEQ ID NO:2. The sialyltransferase-encoding polynucleotide sequences of the invention are typically at least about 75% identical to the nucleic acid sequence of SEQ ID NO:2 over a region at least about 50 nucleotides in length. More preferably, the sialyltransferase-encoding nucleic acids of the invention are at least about 85% identical to this nucleotide sequence, and still more preferably are at least about 95% identical to the nucleotide sequence of SEQ ID NO:2, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of the specified percent identity threshold extends over a longer region than 50 nucleotides, more preferably over a region of at least about 100 nucleotides, and most preferably over the full length of the sialyltransferase-encoding region. Accordingly, the invention provides bifunctional sialyltransferase-encoding nucleic acids that are substantially identical to that of the *C. jejuni* strain OH4384 *cstII* as set forth in SEQ ID NO:2 or strain O:10 (SEQ ID NO:4).

32

Other sialyltransferase-encoding nucleic acids of the invention encode sialyltransferases have α2,3 sialyltransferase activity but lack substantial α2,8 sialyltransferase activity. For example, nucleic acids that encode a *CstII* α2,3 sialyltransferase from *C. jejuni* serostrain O:19 (SEQ ID NO:8) and NCTC 11168 are

5   provided by the invention; these enzymes have little or no α2,8-sialyltransferase activity (Table 6).

    To identify nucleic acids of the invention, one can use visual inspection, or can use a suitable alignment algorithm. An alternative method by which one can identify a bifunctional sialyltransferase-encoding nucleic acid of the invention is by hybridizing, under

10   stringent conditions, the nucleic acid of interest to a nucleic acid that includes a polynucleotide sequence of a sialyltransferase as set forth herein.

2.   *β1,4-GalNAc transferases*

    Also provided by the invention are nucleic acids that include polynucleotide sequences that encode a GalNAc transferase polypeptide that has a β1,4-GalNAc transferase

15   activity. The polynucleotide sequences encode a GalNAc transferase polypeptide that has an amino acid sequence that is at least about 70% identical to the *C. jejuni* OH4384 β1,4-GalNAc transferase, which has an amino acid sequence as set forth in SEQ ID NO:17, over a region at least about 50 amino acids in length. More preferably the GalNAc transferase polypeptide encoded by the nucleic acids of the invention are at least about 80% identical to

20   this amino acid sequence, and still more preferably at least about 90% identical to the amino acid sequence of SEQ ID NO:17, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 50 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the GalNAc transferase polypeptide. In a presently

25   preferred embodiment, the GalNAc transferase polypeptide-encoding nucleic acids of the invention encode a polypeptide having the amino acid sequence as shown in SEQ ID NO:17. To identify nucleic acids of the invention, one can use visual inspection, or can use a suitable alignment algorithm.

    One example of a GalNAc transferase-encoding nucleic acid of the invention

30   is an isolated and/or recombinant form of the GalNAc transferase-encoding nucleic acid of

C. jejuni OH4384. This nucleic acid has a nucleotide sequence as shown in SEQ ID NO:16. The GalNAc transferase-encoding polynucleotide sequences of the invention are typically at least about 75% identical to the nucleic acid sequence of SEQ ID NO:16 over a region at least about 50 nucleotides in length. More preferably, the GalNAc transferase-encoding

5   nucleic acids of the invention are at least about 85% identical to this nucleotide sequence, and still more preferably are at least about 95% identical to the nucleotide sequence of SEQ ID NO:16, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 50 nucleotides, more preferably over a region of at least about 100 nucleotides, and most

10   preferably over the full length of the GalNAc transferase-encoding region.

        To identify nucleic acids of the invention, one can use visual inspection, or can use a suitable alignment algorithm. An alternative method by which one can identify a GalNAc transferase-encoding nucleic acid of the invention is by hybridizing, under stringent conditions, the nucleic acid of interest to a nucleic acid that includes a polynucleotide

15   sequence of SEQ ID NO:16.

3.   *β1,3-Galactosyltransferases*

        The invention also provides nucleic acids that include polynucleotide sequences that encode a polypeptide that has β1,3-galactosyltransferase activity (CgtB). The β1,3-galactosyltransferase polypeptides encoded by these nucleic acids of the invention

20   preferably include an amino acid sequence that is at least about 75% identical to an amino acid sequence of a *C. jejuni* strain OH4384 β1,3-galactosyltransferase as set forth in SEQ ID NO:27, or to that of a strain NCTC 11168 β1,3-galactosyltransferase as set forth in SEQ ID NO:29, over a region at least about 50 amino acids in length. More preferably, the galactosyltransferase polypeptides encoded by these nucleic acids of the invention are at

25   least about 85% identical to this amino acid sequence, and still more preferably are at least about 95% identical to the amino acid sequence of SEQ ID NO:27 or SEQ ID NO:29, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 50 amino acids, more preferably over a region of at least about 100 amino acids, and most preferably over the full length of the

30   galactosyltransferase polypeptide-encoding region.

One example of a β1,3-galactosyltransferase-encoding nucleic acid of the invention is an isolated and/or recombinant form of the β1,3-galactosyltransferase-encoding nucleic acid of *C. jejuni* OH4384. This nucleic acid includes a nucleotide sequence as shown in SEQ ID NO:26. Another suitable β1,3-galactosyltransferase-encoding nucleic acid

5   includes a nucleotide sequence of a *C. jejuni* NCTC 11168 strain, for which the nucleotide sequence is shown in SEQ ID NO:28. The β1,3-galactosyltransferase-encoding polynucleotide sequences of the invention are typically at least about 75% identical to the nucleic acid sequence of SEQ ID NO:26 or that of SEQ ID NO:28 over a region at least about 50 nucleotides in length. More preferably, the β1,3-galactosyltransferase-encoding

10   nucleic acids of the invention are at least about 85% identical to at least one of these nucleotide sequences, and still more preferably are at least about 95% identical to the nucleotide sequences of SEQ ID NO:26 and/or SEQ ID NO:28, over a region of at least 50 amino acids in length. In presently preferred embodiments, the region of percent identity extends over a longer region than 50 nucleotides, more preferably over a region of at least

15   about 100 nucleotides, and most preferably over the full length of the β1,3-galactosyltransferase-encoding region.

To identify nucleic acids of the invention, one can use visual inspection, or can use a suitable alignment algorithm. An alternative method by which one can identify a galactosyltransferase polypeptide-encoding nucleic acid of the invention is by hybridizing,

20   under stringent conditions, the nucleic acid of interest to a nucleic acid that includes a polynucleotide sequence of SEQ ID NO:26 or SEQ ID NO:28.

4. Additional enzymes involved in LOS biosynthetic pathway

Also provided are nucleic acids that encode other enzymes that are involved in the LOS biosynthetic pathway of prokaryotes such as *Campylobacter*. These nucleic acids

25   encode enzymes such as, for example, sialic acid synthase, which is encoded by open reading frame (ORF) 8a of *C. jejuni* strain OH 4384 and by open reading frame 8b of strain NCTC 11168 (*see*, Table 3), another enzyme involved in sialic acid synthesis, which is encoded by ORF 9a of OH 4384 and 9b of NCTC 11168, and a CMP-sialic acid synthetase which is encoded by ORF 10a and 10b of OH 4384 and NCTC 11168, respectively.

35

The invention also provides nucleic acids that encode an acyltransferase that is involved in lipid A biosynthesis. This enzyme is encoded by open reading frame 2a of *C. jejuni* strain OH4384 and by open reading frame 2B of strain NCTC 11168. Nucleic acids that encode an acetyltransferase are also provided; this enzyme is encoded by ORF 11a of

5   strain OH 4384; no homolog is found in the *LOS* biosynthesis locus of strain NCTC 11168.

Also provided are nucleic acids that encode three additional glycosyltransferases. These enzymes are encoded by ORFs 3a, 4a, and 12a of strain OH 4384 and ORFs 3b, 4b, and 12b of strain NH 11168 (Figure 1).

C.   *Expression Cassettes and Expression of the Glycosyltransferases*

10   The present invention also provides expression cassettes, expression vectors, and recombinant host cells that can be used to produce the glycosyltransferases and other enzymes of the invention. A typical expression cassette contains a promoter operably linked to a nucleic acid that encodes the glycosyltransferase or other enzyme of interest. The expression cassettes are typically included on expression vectors that are introduced into

15   suitable host cells, preferably prokaryotic host cells. More than one glycosyltransferase polypeptide can be expressed in a single host cell by placing multiple transcriptional cassettes in a single expression vector, by constructing a gene that encodes a fusion protein consisting of more than one glycosyltransferase, or by utilizing different expression vectors for each glycosyltransferase.

20   In a preferred embodiment, the expression cassettes are useful for expression of the glycosyltransferases in prokaryotic host cells. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (*lac*) promoter

25   systems (Change *et al., Nature* (1977) 198: 1056), the tryptophan (*trp*) promoter system (Goeddel *et al., Nucleic Acids Res.* (1980) 8: 4057), the *tac* promoter (DeBoer, *et al., Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); and the lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake *et al., Nature* (1981) 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes

30   can be used.

Either constitutive or regulated promoters can be used in the present invention. Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the glycosyltransferase polypeptides is induced. High level expression of heterologous proteins slows cell growth in some situations. Regulated promoters especially suitable for use in *E. coli* include the bacteriophage lambda P_L promoter, the hybrid *trp-lac* promoter (Amann *et al.*, Gene (1983) 25: 167; de Boer *et al.*, *Proc. Natl. Acad. Sci. USA* (1983) 80: 21, and the bacteriophage T7 promoter (Studier *et al.*, *J. Mol. Biol.* (1986); Tabor *et al.*, (1985). These promoters and their use are discussed in Sambrook *et al.*, *supra*. A presently preferred regulable promoter is the dual *tac-gal* promoter, which is described in PCT/US97/20528 (Int'l. Publ. No. WO 9820111).

For expression of glycosyltransferase polypeptides in prokaryotic cells other than *E. coli*, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, a hybrid *trp-lac* promoter functions in *Bacillus* in addition to *E. coli*. Promoters suitable for use in eukaryotic host cells are well known to those of skill in the art.

A ribosome binding site (RBS) is conveniently included in the expression cassettes of the invention that are intended for use in prokaryotic host cells. An RBS in *E. coli*, for example, consists of a nucleotide sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine and Dalgarno, *Nature* (1975) 254: 34; Steitz, *In Biological regulation and development: Gene expression* (ed. R.F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

Translational coupling can be used to enhance expression. The strategy uses a short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. *See* Squires *et. al.* (1988) *J. Biol. Chem.* 263: 16297-16302.

The glycosyltransferase polypeptides of the invention can be expressed intracellularly, or can be secreted from the cell. Intracellular expression often results in high yields. If necessary, the amount of soluble, active glycosyltransferase polypeptides can be increased by performing refolding procedures (*see, e.g.*, Sambrook *et al.*, *supra.*; Marston *et*

5   *al.*, *Bio/Technology* (1984) 2: 800; Schoner *et al.*, *Bio/Technology* (1985) 3: 151). In embodiments in which the glycosyltransferase polypeptides are secreted from the cell, either into the periplasm or into the extracellular medium, the polynucleotide sequence that encodes the glycosyltransferase is linked to a polynucleotide sequence that encodes a cleavable signal peptide sequence. The signal sequence directs translocation of the

10   glycosyltransferase polypeptide through the cell membrane. An example of a suitable vector for use in *E. coli* that contains a promoter-signal sequence unit is pTA1529, which has the *E. coli pho*A promoter and signal sequence (*see, e.g.*, Sambrook *et al.*, *supra.*; Oka *et al.*, *Proc. Natl. Acad. Sci. USA* (1985) 82: 7212; Talmadge *et al.*, *Proc. Natl. Acad. Sci. USA* (1980) 77: 3988; Takahara *et al.*, *J. Biol. Chem.* (1985) 260: 2670).

15   The glycosyltransferase polypeptides of the invention can also be produced as fusion proteins. This approach often results in high yields, because normal prokaryotic control sequences direct transcription and translation. In *E. coli*, *lacZ* fusions are often used to express heterologous proteins. Suitable vectors are readily available, such as the pUR, pEX, and pMR100 series (*see, e.g.*, Sambrook *et al.*, *supra.*). For certain applications, it

20   may be desirable to cleave the non-glycosyltransferase amino acids from the fusion protein after purification. This can be accomplished by any of several methods known in the art, including cleavage by cyanogen bromide, a protease, or by Factor X_a (*see, e.g.*, Sambrook *et al.*, *supra.*; Itakura *et al.*, *Science* (1977) 198: 1056; Goeddel *et al.*, *Proc. Natl. Acad. Sci. USA* (1979) 76: 106; Nagai *et al.*, *Nature* (1984) 309: 810; Sung *et al.*, *Proc. Natl. Acad.*

25   *Sci. USA* (1986) 83: 561). Cleavage sites can be engineered into the gene for the fusion protein at the desired point of cleavage.

A suitable system for obtaining recombinant proteins from *E. coli* which maintains the integrity of their N-termini has been described by Miller *et al. Biotechnology* 7:698-704 (1989). In this system, the gene of interest is produced as a C-terminal fusion to

30   the first 76 residues of the yeast ubiquitin gene containing a peptidase cleavage site.

Cleavage at the junction of the two moieties results in production of a protein having an intact authentic N-terminal residue.

Glycosyltransferases of the invention can be expressed in a variety of host cells, including *E. coli,* other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines. Examples of useful bacteria include, but are not limited to, *Escherichia, Enterobacter, Azotobacter, Erwinia, Bacillus, Pseudomonas, Klebsielia, Proteus, Salmonella, Serratia, Shigella, Rhizobia, Vitreoscilla,* and *Paracoccus.* The recombinant glycosyltransferase-encoding nucleic acid is operably linked to appropriate expression control sequences for each host. For *E. coli* this includes a promoter such as the T7, trp, or lambda promoters, a ribosome binding site and preferably a transcription termination signal. For eukaryotic cells, the control sequences will include a promoter and preferably an enhancer derived from immunoglobulin genes, SV40, cytomegalovirus, *etc.,* and a polyadenylation sequence, and may include splice donor and acceptor sequences.

The expression vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells. Cells transformed by the plasmids can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the *amp, gpt, neo* and *hyg* genes.

Once expressed, the recombinant glycosyltransferase polypeptides can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (*see,* generally, R. Scopes, *Protein Purification,* Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol. 182: Guide to Protein Purification.,* Academic Press, Inc. N.Y. (1990)). Substantially pure compositions of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity are most preferred. Once purified, partially or to homogeneity as desired, the polypeptides may then be used (*e.g.,* as immunogens for antibody production). The glycosyltransferases can also be used in an unpurified or semi-purified state. For example, a host cell that expresses the glycosyltransferase can be used

directly in a glycosyltransferase reaction, either with or without processing such as permeabilization or other cellular disruption.

One of skill would recognize that modifications can be made to the glycosyltransferase proteins without diminishing their biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located restriction sites or termination codons or purification sequences.

D.   *Methods and reaction mixtures for synthesis of oligosaccharides*

The invention provides reaction mixtures and methods in which the glycosyltransferases of the invention are used to prepare desired oligosaccharides (which are composed of two or more saccharides). The glycosyltransferase reactions of the invention take place in a reaction medium comprising at least one glycosyltransferase, a donor substrate, an acceptor sugar and typically a soluble divalent metal cation. The methods rely on the use of the glycosyltransferase to catalyze the addition of a saccharide to a substrate (also referred to as an "acceptor") saccharide. A number of methods of using glycosyltransferases to synthesize desired oligosaccharide structures are known. Exemplary methods are described, for instance, WO 96/32491, Ito *et al.* (1993) *Pure Appl. Chem.* 65:753, and U.S. Patents 5,352,670, 5,374,541, and 5,545,553.

For example, the invention provides methods for adding sialic acid in an α2,3 linkage to a galactose residue, by contacting a reaction mixture comprising an activated sialic acid (*e.g.*, CMP-NeuAc, CMP-NeuGc, and the like) to an acceptor moiety that includes a terminal galactose residue in the presence of a bifunctional sialyltransferase of the invention. In presently preferred embodiments, the methods also result in the addition of a second sialic acid residue which is linked to the first sialic acid by an α2,8 linkage. The product of this method is Siaα2,8-Siaα2,3-Gal-. Examples of suitable acceptors include a terminal Gal that is linked to GlcNAc or Glc by a β1,4 linkage, and a terminal Gal that is β1,3-linked to either GlcNAc or GalNAc. The terminal residue to which the sialic acid is attached can itself be attached to, for example, H, a saccharide, oligosaccharide, or an

aglycone group having at least one carbohydrate atom. In some embodiments, the acceptor residue is a portion of an oligosaccharide that is attached to a protein, lipid, or proteoglycan, for example.

In some embodiments, the invention provides reaction mixtures and methods for synthesis of gangliosides, lysogangliosides, ganglioside mimics, lysoganglioside mimics, or the carbohydrate portions of these molecules. These methods and reaction mixtures typically include as the galactosylated acceptor moiety a compound having a formula selected from the group consisting of Gal4Glc-R^1 and Gal3GalNAc-R^2; wherein R^1 is selected from the group consisting of ceramide or other glycolipid, R^2 is selected from the group consisting of Gal4GlcCer, (Neu5Ac3)Gal4GlcCer, and (Neu5Ac8Neu5c3)Gal4GlcCer. For example, for ganglioside synthesis the galactosylated acceptor can be selected from the group consisting of Gal4GlcCer, Gal3GalNAc4(Neu5Ac3)Gal4GlcCer, and Gal3GalNAc4(Neu5Ac8Neu5c3) Gal4GlcCer.

The methods and reaction mixtures of the invention are useful for producing any of a large number of gangliosides, lysogangliosides, and related structures. Many gangliosides of interest are described in Oettgen, H.F., ed., *Gangliosides and Cancer*, VCH, Germany, 1989, pp. 10-15, and references cited therein. Gangliosides of particular interest include, for example, those found in the brain as well as other sources which are listed in Table 1.

Table 1: Ganglioside Formulas and Abbreviations

Structure	Abbreviation
Neu5Ac3Gal4GlcCer	GM3
GalNAc4(Neu5Ac3)Gal4GlcCer	GM2
Gal3GalNAc4(Neu5Ac3)Gal4GlcCer	GM1a
Neu5Ac3Gal3GalNAc4Gal4GlcCer	GM1b
Neu5Ac8Neu5Ac3Gal4GlcCer	GD3
GalNAc4(Neu5Ac8Neu5Ac3)Gal4GlcCer	GD2
Neu5Ac3Gal3GalNAc4(Neu5Ac3)Gal4GlcCer	GD1a
Neu5Ac3Gal3(Neu5Ac6)GalNAc4Gal4GlcCer	GD1α
Gal3GalNAc4(Neu5Ac8Neu5Ac3)Gal4GlcCer	GD1b
Neu5Ac8Neu5Ac3Gal3GalNAc4(Neu5Ac3)Gal4GlcCer	GT1a

41

Structure	Abbreviation
Neu5Ac3Gal3GalNAc4(Neu5Ac8Neu5Ac3)Gal4GlcCer	GT1b
Gal3GalNAc4(Neu5Ac8Neu5Ac8Neu5Ac3)Gal4GlcCer	GT1c
Neu5Ac8Neu5Ac3Gal3GalNAc4(Neu5Ac8Neu5c3)Gal4GlcCer	GQ1b

Nomenclature of Glycolipids, IUPAC-IUB Joint Commission on Biochemical Nomenclature (Recommendations 1997); *Pure Appl. Chem.* (1997) 69: 2475-2487; *Eur. J. Biochem* (1998) 257: 293-298) (www.chem.qmw.ac.uk/iupac/misc/glylp.html).

5   The bifunctional sialyltransferases of the invention are particularly useful for synthesizing the gangliosides GD1a, GD1b, GT1a, GT1b, GT1c, and GQ1b, or the carbohydrate portions of these gangliosides, for example. The structures for these gangliosides, which are shown in Table 1, requires both an α2,3- and an α2,8- sialyltransferase activity. An advantage provided by the methods and reaction mixtures of the invention is that both activities are present in a single polypeptide.

10   The glycosyltransferases of the invention can be used in combination with additional glycosyltransferases and other enzymes. For example, one can use a combination of sialyltransferase and galactosyltransferases. In some embodiments of the invention, the galactosylated acceptor that is utilized by the bifunctional sialyltransferase is formed by contacting a suitable acceptor with UDP-Gal and a galactosyltransferase. The

15   galactosyltransferase polypeptide, which can be one that is described herein, transfers the Gal residue from the UDP-Gal to the acceptor.

Similarly, one can use the β1,4-GalNAc transferases of the invention to synthesize an acceptor for the galactosyltransferase. For example, the acceptor saccharide for the galactosyltransferase can formed by contacting an acceptor for a GalNAc transferase

20   with UDP-GalNAc and a GalNAc transferase polypeptide, wherein the GalNAc transferase polypeptide transfers the GalNAc residue from the UDP-GalNAc to the acceptor for the GalNAc transferase.

In this group of embodiments, the enzymes and substrates can be combined in an initial reaction mixture, or the enzymes and reagents for a second glycosyltransferase

25   cycle can be added to the reaction medium once the first glycosyltransferase cycle has neared completion. By conducting two glycosyltransferase cycles in sequence in a single vessel, overall yields are improved over procedures in which an intermediate species is isolated. Moreover, cleanup and disposal of extra solvents and by-products is reduced.

42

The products produced by the above processes can be used without purification. However, it is usually preferred to recover the product. Standard, well known techniques for recovery of glycosylated saccharides such as thin or thick layer chromatography, or ion exchange chromatography. It is preferred to use membrane filtration,

5   more preferably utilizing a reverse osmotic membrane, or one or more column chromatographic techniques for the recovery.

E.   Uses of Glycoconjugates Produced using Glycosyltransferases and Methods of the Invention

The oligosaccharide compounds that are made using the glycosyltransferases

10   and methods of the invention can be used in a variety of applications, e.g., as antigens, diagnostic reagents, or as therapeutics. Thus, the present invention also provides pharmaceutical compositions which can be used in treating a variety of conditions. The pharmaceutical compositions are comprised of oligosaccharides made according to the methods described above.

15   Pharmaceutical compositions of the invention are suitable for use in a variety of drug delivery systems. Suitable formulations for use in the present invention are found in *Remington's Pharmaceutical Sciences*, Mace Publishing Company, Philadelphia, PA, 17th ed. (1985). For a brief review of methods for drug delivery, *see*, Langer, *Science* 249:1527-1533 (1990).

20   The pharmaceutical compositions are intended for parenteral, intranasal, topical, oral or local administration, such as by aerosol or transdermally, for prophylactic and/or therapeutic treatment. Commonly, the pharmaceutical compositions are administered parenterally, *e.g.*, intravenously. Thus, the invention provides compositions for parenteral administration which comprise the compound dissolved or suspended in an acceptable

25   carrier, preferably an aqueous carrier, *e.g.*, water, buffered water, saline, PBS and the like. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents, detergents and the like.

These compositions may be sterilized by conventional sterilization

30   techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for

43

use as is, or lyophilized, the lyophilized preparation being combined with a sterile aqueous carrier prior to administration. The pH of the preparations typically will be between 3 and 11, more preferably from 5 to 9 and most preferably from 7 and 8.

In some embodiments the oligosaccharides of the invention can be
5   incorporated into liposomes formed from standard vesicle-forming lipids. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al., Ann. Rev. Biophys. Bioeng. 9:467 (1980), U.S. Pat. Nos. 4,235,871, 4,501,728 and 4,837,028. The targeting of liposomes using a variety of targeting agents (e.g., the sialyl galactosides of the invention) is well known in the art (see, e.g., U.S. Patent Nos. 4,957,773 and 4,603,044).
10   The compositions containing the oligosaccharides can be administered for prophylactic and/or therapeutic treatments. In therapeutic applications, compositions are administered to a patient already suffering from a disease, as described above, in an amount sufficient to cure or at least partially arrest the symptoms of the disease and its complications. An amount adequate to accomplish this is defined as a "therapeutically
15   effective dose." Amounts effective for this use will depend on the severity of the disease and the weight and general state of the patient, but generally range from about 0.5 mg to about 40 g of oligosaccharide per day for a 70 kg patient, with dosages of from about 5 mg to about 20 g of the compounds per day being more commonly used.

Single or multiple administrations of the compositions can be carried out with
20   dose levels and pattern being selected by the treating physician. In any event, the pharmaceutical formulations should provide a quantity of the oligosaccharides of this invention sufficient to effectively treat the patient.

The oligosaccharides may also find use as diagnostic reagents. For example, labeled compounds can be used to locate areas of inflammation or tumor metastasis in a
25   patient suspected of having an inflammation. For this use, the compounds can be labeled with appropriate radioisotopes, for example, ^{125}I, ^{14}C, or tritium.

The oligosaccharide of the invention can be used as an immunogen for the production of monoclonal or polyclonal antibodies specifically reactive with the compounds of the invention. The multitude of techniques available to those skilled in the art for
30   production and manipulation of various immunoglobulin molecules can be used in the

44

present invention. Antibodies may be produced by a variety of means well known to those of skill in the art.

The production of non-human monoclonal antibodies, *e.g.*, murine, lagomorpha, equine, *etc.*, is well known and may be accomplished by, for example,
5   immunizing the animal with a preparation containing the oligosaccharide of the invention. Antibody-producing cells obtained from the immunized animals are immortalized and screened, or screened first for the production of the desired antibody and then immortalized. For a discussion of general procedures of monoclonal antibody production, *see*, Harlow and Lane, *Antibodies, A Laboratory Manual* Cold Spring Harbor Publications, N.Y. (1988).
10

EXAMPLE

The following example is offered to illustrate, but not to limit the present invention.

This Example describes the use of two strategies for the cloning of four genes
15   responsible for the biosynthesis of the GT1a ganglioside mimic in the LOS of a bacterial pathogen, *Campylobacter jejuni* OH4384, which has been associated with Guillain-Barré syndrome (Aspinall *et al.* (1994) *Infect. Immun.* 62: 2122-2125). Aspinal *et al.* ((1994) *Biochemistry* 33: 241-249) showed that this strain has an outer core LPS that mimics the tri-sialylated ganglioside GT1a. We first cloned a gene encoding an α-2,3-sialyltransferase (*cst-*
20   *I*) using an activity screening strategy. We then used raw nucleotide sequence information from the recently completed sequence of *C. jejuni* NCTC 11168 to amplify a region involved in LOS biosynthesis from *C. jejuni* OH4384. Using primers that are located in the heptosyl-transferases I and II, the 11.47 kb LOS biosynthesis locus from *C. jejuni* OH4384 was amplified. Sequencing revealed that the locus encodes 13 partial or complete open reading
25   frames (ORFs), while the corresponding locus in *C. jejuni* NCTC 11168 spans 13.49 kb and contains 15 ORFs, indicating a different organization between these two strains.

Potential glycosyltransferase genes were cloned individually, expressed in *Escherichia coli* and assayed using synthetic fluorescent oligosaccharides as acceptors. We identified genes that encode a β-1,4-*N*-acetylgalactosaminyl-transferase (*cgtA*), a β-1,3-
30   galactosyltransferase (*cgtB*) and a bifunctional sialyltransferase (*cst-II*) which transfers sialic

acid to O-3 of galactose and to O-8 of a sialic acid that is linked α-2,3- to a galactose. The linkage specificity of each identified glycosyltransferase was confirmed by NMR analysis at 600 MHz on nanomole amounts of model compounds synthesized *in vitro*. Using a gradient inverse broadband nano-NMR probe, sequence information could be obtained by detection

5   of ^{3}J(C, H) correlations across the glycosidic bond. The role of *cgtA* and *cst-II* in the synthesis of the GT1a mimic in *C. jejuni* OH4384 were confirmed by comparing their sequence and activity with corresponding homologues in two related *C. jejuni* strains that express shorter ganglioside mimics in their LOS. Thus, these three enzymes can be used to synthesize a GT1a mimic starting from lactose.

10

The abbreviations used are: CE, capillary electrophoresis; CMP-Neu5Ac, cytidine monophosphate-*N*-acetylneuraminic acid ; COSY, correlated spectroscopy; FCHASE, 6-(5-fluorescein-carboxamido)-hexanoic acid succimidyl ester; GBS, Guillain-Barré syndrome; HMBC, heteronuclear multiple bond coherence; HSQC, heteronuclear

15   single quantum coherence; LIF, laser induced fluorescence; LOS, lipooligosaccharide; LPS, lipopolysaccharide; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy.

Experimental Procedures

Bacterial strains

20   The following *C. jejuni* strains were used in this study: serostain O:19 (ATCC #43446); serotype O:19 (strains OH4382 and OH4384 were obtained from the Laboratory Centre for Disease Control (Health Canada, Winnipeg, Manitoba)); and serotype O:2 (NCTC #11168). *Escherichia coli* DH5α was used for the *Hind*III library while *E. coli* AD202 (CGSG #7297) was used to express the different cloned glycosyltransferases.

25   ### Basic recombinant DNA methods.

Genomic DNA isolation from the C. jejuni strains was performed using Qiagen Genomic-tip 500/G (Qiagen Inc., Valencia, CA) as described previously (Gilbert *et al.* (1996) *J. Biol. Chem.* 271: 28271-28276). Plasmid DNA isolation, restriction enzyme digestions, purification of DNA fragments for cloning, ligations and transformations were

performed as recommended by the enzyme supplier, or the manufacturer of the kit used for the particular procedure. Long PCR reactions (> 3 kb) were performed using the ExpandTM long template PCR system as described by the manufacturer (Boehringer Mannheim, Montreal). PCR reactions to amplify specific ORFs were performed using the Pwo DNA

5 polymerase as described by the manufacturer (Boehringer Mannheim, Montreal). Restriction and DNA modification enzymes were purchased from New England Biolabs Ltd. (Mississauga, ON). DNA sequencing was performed using an Applied Biosystems (Montreal) model 370A automated DNA sequencer and the manufacturer's cycle sequencing kit.

10 *Activity screening for sialyltransferase from C. jejuni*

The genomic library was prepared using a partial *Hind*III digest of the chromosomal DNA of *C. jejuni* OH4384. The partial digest was purified on a QIAquick column (QIAGEN Inc.) and ligated with *Hind*III digested pBluescript SK-. *E. coli* DH5α was electroporated with the ligation mixture and the cells were plated on LB medium with 150

15 μg/mL ampicillin, 0.05 mM IPTG and 100 μg/mL X-Gal (5-Bromo-4-chloro-indolyl-β-D-galactopyranoside). White colonies were picked in pools of 100 and were resuspended in 1 mL of medium with 15% glycerol. Twenty μL of each pool were used to inoculate 1.5 mL of LB medium supplemented with 150 μg/mL ampicillin. After 2 h of growth at 37 °C, IPTG was added to 1 mM and the cultures were grown for another 4.5 h. The cells were recovered by

20 centrifugation, resuspended in 0.5 mL of 50 mM Mops (pH 7, 10 mM MgCl$_2$) and sonicated for 1 min. The extracts were assayed for sialyltransferase activity as described below except that the incubation time and temperature were 18 h and 32 °C, respectively. The positive pools were plated for single colonies, and 200 colonies were picked and tested for activity in pools of 10. Finally the colonies of the positive pools were tested individually which led to the isolation

25 of a two positive clones, pCJH9 (5.3 kb insert) and pCJH101 (3.9 kb insert). Using several sub-cloned fragments and custom-made primers, the inserts of the two clones were completely sequenced on both strands. The clones with individual *Hind*III fragments were also tested for sialyltransferase activity and the insert of the only positive one (a 1.1 kb Hind*III* fragment cloned in pBluescript SK-) was transferred to pUC118 using *Kpn*I and *Pst*I sites in order to

30 obtain the insert in the opposite orientation with respect to the *plac* promoter.

47

Cloning and sequencing of the LPS biosynthesis locus.

The primers used to amplify the LPS biosynthesis locus of *C. jejuni* OH4384 were based on preliminary sequences available from the website (URL: http://www.sanger.ac.uk/Projects/C_jejuni/) of the *C. jejuni* sequencing group (Sanger

5   Centre, UK) who sequenced the complete genome of the strain NCTC11168. The primers CJ-42 and CJ-43 (all primers sequences are described in Table 2) were used to amplify an 11.47 kb locus using the Expand™ long template PCR system. The PCR product was purified on a S-300 spin column (Pharmacia Biotech) and completely sequence on both strands using a combination of primer walking and sub-cloning of *Hind*III fragments.

10   Specific ORF's were amplified using the primers described in Table 2 and the Pwo DNA polymerase. The PCR products were digested using the appropriate restriction enzymes (*see* Table 2) and were cloned in pCWori+.

Table 2: Primers used for Amplification of Open Reading Frames

Primers used to amplify the LPS core biosynthesis locus

CJ42: Primer in heptosylTase-II **(Error! Reference source not found.)**

5' GC CAT TAC CGT ATC GCC TAA CCA GG 3'   25 mer

CJ43: Primer in heptosylTase-I **(Error! Reference source not found.)**

5' AAA GAA TAC GAA TTT GCT AAA GAG G 3'   25 mer

Primers used to amplify and clone ORF 5a:

CJ-106 (3' primer, 41 mer)(**Error! Reference source not found.**):

            *Sal*I
5' CCT AGG *TCG ACT* TAA AAC AAT GTT AAG AAT ATT TTT TTT AG 3'

CJ-157 (5' primer, 37 mer)(**Error! Reference source not found.**):

            *Nde*I
5' CTT AGG AGG TCA *TAT* GCT ATT TCA ATC ATA CTT TGT G 3'

48

Primers used to amplify and clone ORF 6a:

CJ-105 (3' primer, 37 mer)(**Error! Reference source not found.**):
       *Sal*I
5' CCT AGG *TCG ACC* TCT AAA AAA AAT ATT CTT AAC ATT G 3'

CJ-133 (5' primer, 39 mer)(**Error! Reference source not found.**):
       *Nde*I
5' CTTAGGAGGT*CATA*TGTTTAAAATTTCAATCATCTTACC 3'

Primers used to amplify and clone ORF 7a:

CJ-131 (5' primer, 41 mer)(**Error! Reference source not found.**):

           *Nde*I
5' CTTAGGAGGT*CATA*TGAAAAAAGTTATTATTGCTGGAAATG 3'

CJ-132 (3' primer, 41 mer)(**Error! Reference source not found.**):

      *Sal*I
5' CCTAG*GTCGA*CTTATTTTCCTTTGAAATAATGCTTTATATC 3'

Expression in E. coli *and glycosyltransferase assays.*

The various constructs were transferred to *E. coli* AD202 and were tested for the expression of glycosyltransferase activities following a 4 h induction with 1 mM IPTG. Extracts were made by sonication and the enzymatic reactions were performed overnight at

5   32°C. FCHASE-labeled oligosaccharides were prepared as described previously (Wakarchuk *et al.* (1996) *J. Biol. Chem.* 271: 19166-19173). Protein concentration was determined using the bicinchoninic acid protein assay kit (Pierce, Rockford, IL). For all of the enzymatic assays one unit of activity was defined as the amount of enzyme that generated one μmol of product per minute.

10   The screening assay for α-2,3-sialyltransferase activity in pools of clones contained 1 mM Lac-FCHASE, 0.2 mM CMP-Neu5Ac, 50 mM Mops pH 7, 10 mM MnCl₂ and 10 mM MgCl₂ in a final volume of 10 μL. The various subcloned ORFs were tested for the expression of glycosyltransferase activities following a 4 h induction of the cultures with 1 mM IPTG. Extracts were made by sonication and the enzymatic reactions were performed

15   overnight at 32°C.

The β-1,3-galactosyltransferase was assayed using 0.2 mM GM2-FCHASE, 1 mM UDP-Gal, 50 mM Mes pH 6, 10 mM $MnCl_2$ and 1 mM DTT. The β-1,4-GalNAc transferase was assayed using 0.5 mM GM3-FCHASE, 1mM UDP-GalNAc, 50 mM Hepes pH 7 and 10 mM $MnCl_2$. The α-2,3-sialyltransferase was assayed using 0.5 mM Lac-

5   FCHASE, 0.2 mM CMP-Neu5Ac, 50 mM Hepes pH 7 and 10 mM $MgCl_2$. The α-2,8-sialyltransferase was assayed using 0.5 mM GM3-FCHASE, 0.2 mM CMP-Neu5Ac, 50 mM Hepes pH 7 and 10 mM $MnCl_2$.

The reaction mixes were diluted appropriately with 10 mM NaOH and analyzed by capillary electrophoresis performed using the separation and detection conditions

10   as described previously (Gilbert *et al.* (1996) *J. Biol. Chem.* 271, 28271-28276). The peaks from the electropherograms were analyzed using manual peak integration with the P/ACE Station software. For rapid detection of enzyme activity, samples from the transferase reaction mixtures were examined by thin layer chromatography on silica-60 TLC plates (E. Merck) as described previously (*Id.*).

15   *NMR spectroscopy*

NMR experiments were performed on a Varian INOVA 600 NMR spectrometer. Most experiments were done using a 5 mm Z gradient triple resonance probe. NMR samples were prepared from 0.3-0.5 mg (200-500 nanomole) of FCHASE-glycoside. The compounds were dissolved in H_2O and the pH was adjusted to 7.0 with dilute NaOH.

20   After freeze drying the samples were dissolved in 600 µL D_2O. All NMR experiments were performed as previously described (Pavliak *et al.* (1993) *J. Biol. Chem.* 268: 14146-14152; Brisson *et al.* (1997) *Biochemistry* 36: 3278-3292) using standard techniques such as COSY, TOCSY, NOESY, 1D-NOESY, 1D-TOCSY and HSQC. For the proton chemical shift reference, the methyl resonance of internal acetone was set at 2.225 ppm (^1H). For the

25   ^{13}C chemical shift reference, the methyl resonance of internal acetone was set at 31.07 ppm relative to external dioxane at 67.40 ppm. Homonuclear experiments were on the order of 5-8 hours each. The 1D NOESY experiments for GD3-FCHASE ,[0.3 mM], with 8000 scans and a mixing time of 800 ms was done for a duration of 8.5 h each and processed with a line broadening factor of 2-5 Hz. For the 1D NOESY of the resonances at 4.16 ppm, 3000 scans

30   were used. The following parameters were used to acquire the HSQC spectrum: relaxation

delay of 1.0 s, spectral widths in F_2 and F_1 of 6000 and 24147 Hz, respectively, acquisition times in t_2 of 171 ms. For the t_1 dimension, 128 complex points were acquired using 256 scans per increment. The sign discrimination in F_1 was achieved by the States method. The total acquisition time was 20 hours. For GM2-FCHASE, due to broad lines, the number of scans per increment was increased so that the HSQC was performed for 64 hours. The phase-sensitive spectrum was obtained after zero filling to 2048 x 2048 points. Unshifted gaussian window functions were applied in both dimensions. The HSQC spectra were plotted at a resolution of 23 Hz / point in the ^{13}C dimension and 8 Hz/ point in the proton dimension. For the observation of the multiplet splittings, the ^{1}H dimension was reprocessed at a resolution of 2 Hz / point using forward linear prediction and a π/4-shifted squared sinebell function. All the NMR data was acquired using Varian's standard sequences provided with the VNMR 5.1 or VNMR 6.1 software. The same program was used for processing.

A gradient inverse broadband nano-NMR probe (Varian) was used to perform the gradient HMBC (Bax and Summers (1986) *J. Am. Chem. Soc.* **108**, 2093-2094; Parella *et al.* (1995) *J. Mag. Reson. A* **112**, 241-245) experiment for the GD3-FCHASE sample. The nano-NMR probe which is a high-resolution magic angle spinning probe produces high resolution spectra of liquid samples dissolved in only 40 μL (Manzi *et al.* (1995) *J. Biol. Chem.* **270**, 9154-9163). The GD3-FCHASE sample (mass = 1486.33 Da) was prepared by lyophilizing the original 0.6 mL sample (200 nanomoles) and dissolving it in 40 μL of D_2O for a final concentration of 5 mM. The final pH of the sample could not be measured.

The gradient HMBC experiment was done at a spin rate of 2990 Hz, 400 increments of 1024 complex points, 128 scans per increment, acquisition time of 0.21 s, ^{1}J(C, H) = 140 Hz and ^{n}J(C, H) = 8 Hz, for a duration of 18.5 h.

Mass spectrometry

All mass measurements were obtained using a Perkin-Elmer Biosystems (Fragmingham, MA) Elite-STR MALDI-TOF instrument. Approximately two μg of each oligosaccharide was mixed with a matrix containing a saturated solution of dihydroxybenzoic acid. Positive and negative mass spectra were acquired using the reflector mode.

RESULTS

Detection of glycosyltransferase activities in C. jejuni strains

Before the cloning of the glycosyltransferase genes, we examined *C. jejuni*
OH4384 and NCTC 11168 cells for various enzymatic activities. When an enzyme activity
5   was detected, the assay conditions were optimized (described in the Experimental
Procedures) to ensure maximal activity. The capillary electrophoresis assay we employed
was extremely sensitive and allowed detection of enzyme activity in the μU/ml range
(Gilbert *et al*. (1996) *J. Biol. Chem.* 271: 28271-28276). We examined both the sequenced
strain NCTC 11168 and the GBS-associated strain OH4384 for the enzymes required for the
10 GT1a ganglioside mimic synthesis. As predicted, strain OH4384 possessed the enzyme
activities required for the synthesis of this structure: β-1,4-N-
acetylgalactosaminyltransferase, β-1,3-galactosyltransferase, α-2,3-sialyltransferase and α-
2,8-sialyltransferase. The genome of the strain, NCTC 11168 lacked the β-1,3-
galactosyltransferase and the α-2,8-sialyltransferase activities.

15   ### *Cloning of an α-2,3-sialyltransferase (cst-I) using an activity screening strategy*

A plasmid library made from an unfractionated partial *Hind*III digestion of
chromosomal DNA from *C. jejuni* OH4384 yielded 2,600 white colonies which were picked
to form pools of 100. We used a "divide and conquer" screening protocol from which two
positive clones were obtained and designated pCJH9 (5.3 kb insert, 3 *Hind*III sites) and
20   pCJH101 (3.9 kb insert, 4 *Hind*III sites). Open reading frame (ORF) analysis and PCR
reactions with *C. jejuni* OH4384 chromosomal DNA indicated that pCJH9 contained inserts
that were not contiguous in the chromosomal DNA. The sequence downstream of nucleotide
#1440 in pCJH9 was not further studied while the first 1439 nucleotides were found to be
completely contained within the sequence of pCJH101. The ORF analysis and PCR reactions
25   with chromosomal DNA indicated that all of the pCJH101 *Hind*III fragments were
contiguous in *C. jejuni* OH4384 chromosomal DNA.

Four ORFs, two partial and two complete, were found in the sequence of
pCJH101 (Figure 2). The first 812 nucleotides encode a polypeptide that is 69 % identical
with the last 265 a.a. residues of the peptide chain release factor RF-2 (*prfB* gene, GenBank
30   #AE000537) from *Helicobacter pylori*. The last base of the TAA stop codon of the chain

52

release factor is also the first base of the ATG start codon of an open reading frame that spans nucleotides #812 to #2104 in pCJH101. This ORF was designated *cst-I* (Campylobacter sialyltransferase I) and encodes a 430 amino acid polypeptide that is homologous with a putative ORF from *Haemophilus influenzae* (GenBank #U32720). The

5   putative *H. influenzae* ORF encodes a 231 amino acid polypeptide that is 39 % identical to the middle region of the Cst I polypeptide (amino acid residues #80 to #330). The sequence downstream of *cst-I* includes an ORF and a partial ORF that encode polypeptides that are homologous (> 60 % identical) with the two subunits, CysD and CysN, of the *E. coli* sulfate adenylyltransferase (GenBank #AE000358).

10   In order to confirm that the *cst-I* ORF encodes sialyltransferase activity, we sub-cloned it and over-expressed it in *E. coli*. The expressed enzyme was used to add sialic acid to Gal-β-1,4-Glc-β-FCHASE (Lac-FCHASE). This product (GM3-FCHASE) was analyzed by NMR to confirm the Neu5Ac-α-2,3-Gal linkage specificity of Cst-I.

Sequencing of the LOS biosynthesis locus of C. jejuni OH4384

15   Analysis of the preliminary sequence data available at the website of the *C. jejuni* NCTC 11168 sequencing group (Sanger Centre, UK (http://www.sanger.ac.uk/ Projects /C_jejuni/)) revealed that the two heptosyltransferases involved in the synthesis of the inner core of the LPS were readily identifiable by sequence homology with other bacterial heptosyltransferases. The region between the two heptosyltransferases spans 13.49

20   kb in NCTC 11168 and includes at least seven potential glycosyltransferases based on BLAST searches in GenBank. Since no structure is available for the LOS outer core of NCTC 11168, it was impossible to suggest functions for the putative glycosyltransferase genes in that strain.

Based on conserved regions in the heptosyltransferases sequences, we

25   designed primers (CJ-42 and CJ-43) to amplify the region between them. We obtained a PCR product of 13.49 kb using chromosomal DNA from *C. jejuni* NCTC 11168 and a PCR product of 11.47 kb using chromosomal DNA from *C. jejuni* OH4384. The size of the PCR product from strain NCTC 11168 was consistent with the Sanger Centre data. The smaller size of the PCR product from strain OH4384 indicated heterogeneity between the strains in

30   the region between the two heptosyltransferase genes and suggested that the genes for some

53

of the glycosyltransferases specific to strain OH4384 could be present in that location. We sequenced the 11.47 kb PCR product using a combination of primer walking and sub-cloning of HindIII fragments (GenBank #AF130984). The G/C content of the DNA was 27%, typical of DNA from *Campylobacter*. Analysis of the sequence revealed eleven complete ORFs in addition to the two partial ORFs encoding the two heptosyltransferases (Figure 2, Table 3). When comparing the deduced amino acid sequences, we found that the two strains share six genes that are above 80% identical and four genes that are between 52 and 68% identical (Table 3). Four genes are unique to *C. jejuni* NCTC 11168 while one gene is unique to *C. jejuni* OH4384 (Figure 2). Two genes that are present as separate ORFs (ORF #5a and #10a) in *C. jejuni* OH4384 are found in an in-frame fusion ORF (#5b/10b) in *C. jejuni* NCTC 11168.

Table 3

Location and description of the ORFs of the LOS biosynthesis locus from C. jejuni OH4384

ORF #	Location	Homologue in Strain NCTC11168 [a] (% identity in the a.a. sequence)	Homologues found in GenBank (% identity in the a.a sequence)	Function[b]
1a	1-357	ORF #1b (98%)	*rfaC* (GB #AE000546) from *Helicobacter pylori* (35%)	Heptosyltransferase I
2a	350-1,234	ORF #2b (96%)	*waaM* (GB #AE001463) from *Helicobacter pylori* (25%)	Lipid A biosynthesis acyltransferase
3a	1,234-2,487	ORF #3b (90%)	*lgtF* (GB #U58765) from *Neisseria meningitidis* (31%)	Glycosyltransferase

ORF #	Location	Homologue in Strain NCTC11168[a] (% identity in the a.a. sequence)	Homologues found in GenBank (% identity in the a.a sequence)	Function[b]
4a	2,786-3,952	ORF #4b (80%)	*cps14J* (GB #X85787) from *Streptococcus pneumoniae* (45% over first 100 a.a)	Glycosyltransferase
5a	4,025-5,065	N-terminus of ORF #5b/10b (52%)	ORF #HP0217 (GB #AE000541) from *Helicobacter pylori* (50%)	β-1,4-*N*-acetylgalac-tosaminyltransferase (*cgtA*)
6a	5,057-5,959 (comple ment)	ORF #6b (60%)	*cps23FU* (GB #AF030373) from *Streptococcus pneumoniae (23%)*	β-1,3-Galactosyl-transferase (*cgtB*)
7a	6,048-6,920	ORF #7b (52%)	ORF #HI0352 (GB #U32720) from *Haemophilus influenzae* (40%)	Bi-functional α-2,3/α2,8 sialyl-transferase (*cst-II*)
8a	6,924-7,961	ORF #8b (80%)	*siaC* (GB #U40740) from *Neisseria meningitidis* (56%)	Sialic acid synthase
9a	8,021-9,076	ORF #9b (80%)	*siaA* (GB #M95053) from *Neisseria meningitidis* (40%)	Sialic acid biosynthesis
10a	9,076-9,738	C-terminus of ORF #5b/10b (68%)	*neuA* (GB #U54496) from *Haemophilus ducreyi* (39%)	CMP-sialic acid synthetase

ORF #	Location	Homologue in Strain NCTC11168[a] (% identity in the a.a. sequence)	Homologues found in GenBank (% identity in the a.a sequence)	Function[b]
11a	9,729-10,559	No homologue	Putative ORF (GB #AF010496) from *Rhodobacter capsulatus* (22%)	Acetyltransferase
12a	10,557-11,366 (complement)	ORF #12b (90%)	ORF #HI0868 (GB #U32768) from *Haemophilus influenzae* (23%)	Glycosyltransferase
13a	11,347-11,474	ORF #13b (100%)	*rfaF* (GB #AE000625) from *Helicobacter pylori* (60%)	Heptosyltransferase II

[a] The sequence of the *C. jejuni* NCTC 11168 ORFs can be obtained from the Sanger Centre (URL:http//www.sanger.ac.uk/Projects/C_jejuni/).

[b] The functions that were determined experimentally are in bold fonts. Other functions are based on higher score homologues from GenBank.

Identification of outer core glycosyltransferases

Various constructs were made to express each of the potential glycosyltransferase genes located between the two heptosyltransferases from *C. jejuni* OH4384. The plasmid pCJL-09 contained the ORF #5a and a culture of this construct showed GalNAc transferase activity when assayed using GM3-FCHASE as acceptor. The GalNAc transferase was specific for a sialylated acceptor since Lac-FCHASE was a poor substrate (less than 2% of the activity observed with GM3-FCHASE). The reaction product obtained from GM3-FCHASE had the correct mass as determined by MALDI-TOF mass spectrometry, and the identical elution time in the CE assay as the GM2-FCHASE standard.

Considering the structure of the outer core LPS of *C. jejuni* OH4384, this GalNAc

transferase (*cgtA* for Camplyobacter glycosyltransferase A), has a β-1,4-specificity to the

terminal Gal residue of GM3-FCHASE. The linkage specificity of CgtA was confirmed by

the NMR analysis of GM2-FCHASE (see text below, Table 4). The *in vivo* role of *cgtA* in

5   the synthesis of a GM2 mimic is confirmed by the natural knock-out mutant provided by *C.

jejuni* OH4382 (Figure 1). Upon sequencing of the *cgtA* homologue from *C. jejuni* OH4382

we found a frame-shift mutation (a stretch of seven A's instead of 8 A's after base #71)

which would result in the expression of a truncated cgtA version (29 aa instead of 347 aa).

The LOS outer core structure of *C. jejuni* OH4382 is consistent with the absence of β-1,4-

10   GlaNAc transferase as the inner galactose residue is substituted with sialic acid only

(Aspinall *et al.* (1994) *Biochemistry* **33**, 241-249).

The plasmid pCJL-04 contained the ORF #6a and an IPTG-induced culture of

this construct showed galactosyltransferase activity using GM2-FCHASE as an acceptor

thereby producing GM1a-FCHASE. This product was sensitive to β-1,3-galactosidase and

15   was found to have the correct mass by MALDI-TOF mass spectrometry. Considering the

structure of the LOS outer core of *C. jejuni* OH4384, we suggest that this

galactosyltransferase (*cgtB* for Campylobacter glycosyltransferase **B**) has β-1,3- specificity

to the terminal GalNAc residue of GM2-FCHASE. The linkage specificity of CgtA was

confirmed by the NMR analysis of GM1a-FCHASE (see text below, Table 4) which was

20   synthesized by using sequentially Cst-I, CgtA and CgtB.

The plasmid pCJL-03 included the ORF #7a and an IPTG-induced culture

showed sialyltransferase activity using both Lac-FCHASE and GM3-FCHASE as acceptors.

This second sialyltransferase from OH4384 was designated *cst-II*. Cst-II was shown to be

bi-functional as it could transfer sialic acid α-2,3 to the terminal Gal of Lac-FCHASE and

25   also α-2,8- to the terminal sialic acid of GM3-FCHASE. NMR analysis of a reaction

product formed with Lac-FCHASE confirmed the α-2,3-linkage of the first sialic acid on the

Gal, and the α-2,8-linkage of the second sialic acid (see text below, Table 4).

Table 4

Proton NMR chemical shifts[a] for the fluorescent derivatives of the ganglioside mimics synthesized using the cloned glycosyltransferases.

Residue	H	Chemical Shift (ppm)				
		Lac-	GM3-	GM2-	GM1a-	GD3-
βGlc	1	4.57	4.70	4.73	4.76	4.76
a	2	3.23	3.32	3.27	3.30	3.38
	3	3.47	3.54	3.56	3.58	3.57
	4	3.37	3.48	3.39	3.43	3.56
	5	3.30	3.44	3.44	3.46	3.50
	6	3.73	3.81	3.80	3.81	3.85
	6'	3.22	3.38	3.26	3.35	3.50
βGal(1-4)	1	4.32	4.43	4.42	4.44	4.46
b	2	3.59	3.60	3.39	3.39	3.60
	3	3.69	4.13	4.18	4.18	4.10
	4	3.97	3.99	4.17	4.17	4.00
	5	3.81	3.77	3.84	3.83	3.78
	6	3.86	3.81	3.79	3.78	3.78
	6'	3.81	3.78	3.79	3.78	3.78
αNeu5Ac(2-3)	3_{ax}		1.81	1.97	1.96	1.78
c	3_{eq}		2.76	2.67	2.68	2.67
	4		3.69	3.78	3.79	3.60
	5		3.86	3.84	3.83	3.82
	6		3.65	3.49	3.51	3.68
	7		3.59	3.61	3.60	3.87
	8		3.91	3.77	3.77	4.15
	9		3.88	3.90	3.89	4.18
	9'		3.65	3.63	3.64	3.74
	NAc		2.03	2.04	2.03	2.07
βGalNAc(1-4)	1			4.77	4.81	
d	2			3.94	4.07	
	3			3.70	3.82	
	4			3.93	4.18	

		Chemical Shift (ppm)				
Residue	H	Lac-	GM3-	GM2-	GM1a-	GD3-
	5			3.74	3.75	
	6			3.86	3.84	
	6'			3.86	3.84	
	NAc			2.04	2.04	
βGal(1-3)	1				4.55	
e	2				3.53	
	3				3.64	
	4				3.92	
	5				3.69	
	6				3.78	
	6'				3.74	
αNeu5Ac(2-8)	3_{ax}					1.75
f	3_{eq}					2.76
	4					3.66
	5					3.82
	6					3.61
	7					3.58
	8					3.91
	9					3.88
	9'					3.64
	NAc					2.02

[a] in ppm from HSQC spectrum obtained at 600 MHz, D_2O, pH 7, 28°C for Lac-, 25°C for GM3-, 16°C for GM2-, 24°C for GM1a-, and 24°C GD3-FCHASE. The methyl resonance of internal acetone is at 2.225 ppm (^1H). The error is ± 0.02 ppm for ^1H chemical shifts and ± 5°C for the sample temperature. The error is ± 0.1 ppm for the H-6 resonances of residue a, b, d and e due to overlap.

Comparison of the sialyltransferases

The *in vivo* role of *cst-II* from *C. jejuni* OH4384 in the synthesis of a tri-sialylated GT1a ganglioside mimic is supported by comparison with the *cst-II* homologue from *C. jejuni* O:19 (serostrain) that expresses the di-sialylated GD1a ganglioside mimic. There are 24 nucleotide differences that translate into 8 amino acid differences between

59

these two *cst-II* homologues (Figure 3). When expressed in *E. coli*, the *cst-II* homologue from *C. jejuni* O:19 (serostrain) has α-2,3-sialyltransferase activity but very low α-2,8-sialyltransferase activity (Table 5) which is consistent with the absence of terminal α-2,8-linked sialic acid in the LOS outer core (Aspinall *et al.* (1994) *Biochemistry* 33, 241-249) of

5   *C. jejuni* O:19 (serostrain). The *cst-II* homologue from *C. jejuni* NCTC 11168 expressed much lower α-2,3-sialyltransferase activity than the homologues from O:19 (serostrain) or OH4384 and no detectable α-2,8-sialyltransferase activity. We could detect an IPTG-inducible band on a SDS-PAGE gel when *cst-II* from NCTC 11168 was expressed in *E. coli* (data not shown). The Cst-II protein from NCTC 11168 shares only 52% identity with the

10   homologues from O:19 (serostrain) or OH4384. We could not determine whether the sequence differences could be responsible for the lower activity expressed in *E. coli*.

Although *cst-I* mapped outside the LOS biosynthesis locus, it is obviously homologous to *cst-II* since its first 300 residues share 44% identity with Cst-II from either *C. jejuni* OH4384 or *C. jejuni* NCTC 11168 (Figure 3). The two Cst-II homologues share

15   52% identical residues between themselves and are missing the C-terminal 130 amino acids of Cst-I. A truncated version of Cst-I which was missing 102 amino acids at the C-terminus was found to be active (data not shown) which indicates that the C-terminal domain of Cst-I is not necessary for sialyltransferase activity. Although the 102 residues at the C-terminus are dispensable for *in vitro* enzymatic activity, they may interact with other cell components

20   *in vivo* either for regulatory purposes or for proper cell localization. The low level of conservation between the *C. jejuni* sialyltransferases is very different from what was previously observed for the α-2,3-sialyltransferases from *N. meningitidis* and *N. gonorrhoeae*, where the *lst* transferases are more than 90% identical at the protein level between the two species and between different isolates of the same species (Gilbert *et al.*,

25   *supra.*).

Table 5

Comparison of the activity of the sialyltransferases from C. jejuni. The various sialyltransferases were expressed in *E. coli* as fusion proteins with the maltose-binding

30   protein in the vector pCWori+ (Wakarchuk *et al.* (1994) *Protein. Sci.* 3, 467-475). Sonicated extracts were assayed using 500 μM of either Lac-FCHASE or GM3-FCHASE.

| Sialyltransferase gene | Activity (μU/mg)[a] | | Ratio (%)[b] |
	Lac-FCHASE	GM3-FCHASE	
cst-I (OH4384)	3,744	2.2	0.1
cst-II (OH4384)	209	350.0	167.0
cst-II (O:19 serostrain)	2,084	1.5	0.1
cst-II (NCTC 11168)	8	0	0.0

[a] The activity is expressed in μU (pmol of product per minute) per mg of total protein in the extract.

[b] Ratio (in percentage) of the activity on GM3-FCHASE divided by the activity on Lac-FCHASE.

5   ***NMR analysis on nanomole amounts of the synthesized model compounds.***

In order to properly assess the linkage specificity of an identified glycosyltransferase, its product was analyzed by NMR spectroscopy. In order to reduce the time needed for the purification of the enzymatic products, NMR analysis was conducted on nanomole amounts. All compounds are soluble and give sharp resonances with linewidths of

10   a few Hz since the H-1 anomeric doublets ($J_{1,2}$ = 8 Hz) are well resolved. The only exception is for GM2-FCHASE which has broad lines (~ 10 Hz), probably due to aggregation. For the proton spectrum of the 5 mM GD3-FCHASE solution in the nano-NMR probe, the linewidths of the anomeric signals were on the order of 4 Hz, due to the increased concentration. Also, additional peaks were observed, probably due to degradation of the

15   sample with time. There were also some slight chemical shifts changes, probably due to a change in pH upon concentrating the sample from 0.3 mM to 5 mM. Proton spectra were acquired at various temperatures in order to avoid overlap of the HDO resonance with the anomeric resonances. As can be assessed from the proton spectra, all compounds were pure and impurities or degradation products that were present did not interfere with the NMR

20   analysis which was performed as previously described (Pavliak *et al.* (1993) *J. Biol. Chem.* **268**, 14146-14152; Brisson *et al.* (1997) *Biochemistry* **36**, 3278-3292).

For all of FCHASE glycosides, the ^{13}C assignments of similar glycosides (Sabesan and Paulson (1986) *J. Am. Chem. Soc.* **108**, 2068-2080; Michon *et al.* (1987) *Biochemistry* **26**, 8399-8405; Sabesan *et al.* (1984) *Can. J. Chem.* **62**, 1034-1045) were available. For the FCHASE glycosides, the ^{13}C assignments were verified by first assigning

5 the proton spectrum from standard homonuclear 2D experiments, COSY, TOCSY and NOESY, and then verifying the ^{13}C assignments from an HSQC experiment, which detects C-H correlations. The HSQC experiment does not detect quaternary carbons like C-1 and C-2 of sialic acid, but the HMBC experiment does. Mainly for the Glc resonances, the proton chemical shifts obtained from the HSQC spectra differed from those obtained from

10 homonuclear experiments due to heating of the sample during ^{13}C decoupling. From a series of proton spectrum acquired at different temperatures, the chemical shifts of the Glc residue were found to be the most sensitive to temperature. In all compounds, the H-1 and H-2 resonances of Glc changed by 0.004 ppm / °C, the Gal(1-4) H-1 by 0.002 ppm / °C, and less than 0.001 ppm / °C for the Neu5Ac H-3 and other anomeric resonances. For LAC-

15 FCHASE, the Glc H-6 resonance changed by 0.008 ppm / °C.

The large temperature coefficient for the Glc resonances is attributed to ring current shifts induced by the linkage to the aminophenyl group of FCHASE. The temperature of the sample during the HSQC experiment was measured from the chemical shift of the Glc H-1 and H-2 resonances. For GM1a-FCHASE, the temperature changed

20 from 12°C to 24°C due to the presence of the Na+ counterion in the solution and NaOH used to adjust the pH. Other samples had less severe heating (< 5°C). In all cases, changes of proton chemical shifts with temperature did not cause any problems in the assignments of the resonances in the HSQC spectrum. In Table 4 and Table 6, all the chemical shifts are taken from the HSQC spectra.

25 The linkage site on the aglycon was determined mainly from a comparison of the ^{13}C chemical shifts of the enzymatic product with those of the precursor to determine glycosidation shifts as done previously for ten sialyloligosaccharides (Salloway *et al.* (1996) *Infect. Immun.* **64**, 2945-2949). Here, instead of comparing ^{13}C spectra, HSQC spectra are compared, since one hundred times more material would be needed to obtain a ^{13}C spectrum.

30 When the ^{13}C chemical shifts from HSQC spectra of the precursor compound are compared

to those of the enzymatic product, the main downfield shift always occurs at the linkage site while other chemical shifts of the precursor do not change substantially. Proton chemical shift differences are much more susceptible to long-range conformational effects, sample preparation, and temperature. The identity of the new sugar added can quickly be identified

5 from a comparison of its ^{13}C chemical shifts with those of monosaccharides or any terminal residue, since only the anomeric chemical shift of the glycon changes substantially upon glycosidation (Sabesan and Paulson, *supra.*).

Vicinal proton spin-spin coupling (J_{HH}) obtained from 1D TOCSY or 1D NOESY experiments also are used to determine the identity of the sugar. NOE experiments

10 are done to sequence the sugars by the observation of NOEs between the anomeric glycon protons (H-3s for sialic acid) and the aglycon proton resonances. The largest NOE is usually on the linkage proton but other NOEs can also occur on aglycon proton resonances that are next to the linkage site. Although at 600 MHz, the NOEs of many tetra- and pentasaccharides are positive or very small, all these compounds gave good negative NOEs

15 with a mixing time of 800 ms, probably due to the presence of the large FCHASE moiety.

For the synthetic Lac-FCHASE, the ^{13}C assignments for the lactose moiety of Lac-FCHASE were confirmed by the 2D methods outlined above. All the proton resonances of the Glc unit were assigned from a 1D-TOCSY experiment on the H-1 resonance of Glc with a mixing time of 180 ms. A 1D-TOCSY experiment for Gal H-1 was used to assign the

20 H-1 to H-4 resonances of the Gal unit. The remaining H-5 and H-6s of the Gal unit were then assigned from the HSQC experiment. Vicinal spin-spin coupling values (J_{HH}) for the sugar units were in accord with previous data (Michon *et al.*, *supra.*). The chemical shifts for the FCHASE moiety have been given previously (Gilbert *et al.* (1996) *J. Biol. Chem.* **271**, 28271-28276).

25 Accurate mass determination of the enzymatic product of Cst-I from Lac-FCHASE was consistent with the addition of sialic acid to the Lac-FCHASE acceptor (Figure 4). The product was identified as GM3-FCHASE since the proton spectrum and ^{13}C chemical shifts of the sugar moiety of the product (Table 6) were very similar to those for the GM3 oligosaccharide or sialyllactose, (αNeu5Ac(2-3)βGal(1-4)βGlc; Sabesan and

30 Paulson, *supra.*). The proton resonances of GM3-FCHASE were assigned from the COSY

spectrum, the HSQC spectrum, and comparison of the proton and ^{13}C chemical shifts with those of αNeu5Ac(2-3)βGal(1-4)βGlcNAc-FCHASE (Gilbert *et al.*, *supra.*). For these two compounds, the proton and ^{13}C chemical shifts for the Neu5Ac and Gal residues were within error bounds of each other (*Id.*). From a comparison of the HSQC spectra of Lac-FCHASE

5   and GM3-FCHASE, it is obvious that the linkage site is at Gal C-3 due to the large downfield shift for Gal H-3 and Gal C-3 upon sialylation typical for (2-3) sialyloligosaccharides (Sabesan and Paulson, *supra.*). Also, as seen before for αNeu5Ac(2-3)βGal(1-4)βGlcNAc-FCHASE (Gilbert *et al.*, *supra.*), the NOE from H-3$_{ax}$ of sialic acid to H-3 of Gal was observed typical of the αNeu5Ac(2-3)Gal linkage.

10

Table 6

Comparison of the ^{13}C chemical shifts for the FCHASE glycosides[a] with those observed for lactose[b] (Sabesan and Paulson, supra.), ganglioside oligosaccharides[b] (Id., Sabesan et al.

15   *(1984) Can. J. Chem. 62, 1034-1045) and (-8NeuAc2-)₃ (Michon et al. (1987) Biochemistry* *26, 8399-8405). The chemical shifts at the glycosidation sites are underlined.*

Residue	C	Lac-	Lactose	GM3-	GM3OS	GM2-	GM2OS	GM1a-	GM1a OS	GD3-	8NeuAc2
βGlc	1	100.3	**96.7**	100.3	**96.8**	100.1	**96.6**	100.4	**96.6**	100.6	
a	2	73.5	**74.8**	73.4	**74.9**	73.3	**74.6**	73.3	**74.6**	73.5	
	3	75.2	75.3	75.0	75.4	75.3	75.2	75.0	75.2	75.0	
	4	<u>79.4</u>	<u>79.4</u>	<u>79.0</u>	<u>79.4</u>	<u>79.5</u>	<u>79.5</u>	<u>79.5</u>	<u>79.5</u>	<u>78.8</u>	
	5	75.9	75.7	75.7	75.8	75.8	75.6	75.7	75.6	75.8	
	6	61.1	61.1	60.8	61.2	61.0	61.0	60.6	61.0	60.8	
βGal(1-4)	1	104.1	103.8	103.6	103.7	103.6	103.5	103.6	103.5	103.6	
b	2	72.0	71.9	70.3	70.4	71.0	70.9	70.9	70.9	70.3	
	3	73.5	73.5	<u>76.4</u>	<u>76.6</u>	<u>75.3</u>	<u>75.6[c]</u>	<u>75.1</u>	<u>75.2[c]</u>	76.3	
	4	69.7	69.5	68.4	68.5	<u>78.3</u>	<u>78.0[c]</u>	<u>78.1</u>	<u>78.0[c]</u>	68.5	
	5	76.4	76.3	76.0	76.2	75.0	74.9	74.9	75.0	76.1	
	6	62.1	62.0	62.1	62.0	62.2	61.4	62.0	61.5	62.0	
αNeu5Ac (2-3)	3			40.4	40.7	37.7	37.9	37.8	37.9	40.4	**41.7**
c	4			69.2	69.3	69.8	69.5	69.5	69.5	69.0	68.8[d]
	5			52.6	52.7	52.7	52.5	52.6	52.5	53.0	53.2

64

				Chemical Shift (ppm)							
Residue	C	Lac-	Lactose	GM3-	GM3OS	GM2-	GM2OS	GM1a-	GM1aOS	GD3-	8NeuAc2
	6			73.7	73.9	74.0	73.9	73.8	73.9	74.9	74.5[d]
	7	.		69.0	69.2	69.0	68.8	69.0	68.9	70.3	70.0
	8			72.6	72.8	73.3	73.1	73.1	73.1	79.1	79.1
	9			63.4	63.7	63.9	63.7	63.7	63.7	62.5	62.1
	NAc			22.9	23.1	23.2	22.9	23.3	22.9	23.2	23.2
βGalNAc (1-4) d	1					103.8	103.6	103.4	103.4		
	2					53.2	53.2	52.0	52.0		
	3					72.3	72.2	81.4	81.2		
	4					68.8	68.7	68.9	68.8		
	5					75.6	75.2	75.1	75.2		
	6					61.8	62.0	61.5	62.0		
	NAc					23.2	23.5	23.4	23.5		
βGal(1-3) e	1							105.5	105.6		
	2							71.5	71.6		
	3							73.1	73.4		
	4							69.5	69.5		
	5							75.7	75.8		
	6							61.9	61.8		
αNeu5Ac (2-8) f	3									41.2	41.2
	4									69.5	69.3
	5									53.0	52.6
	6									73.6	73.5
	7									69.0	69.0
	8									72.7	72.6
	9									63.5	63.4
	NAc									23.0	23.1

[a] in ppm from the HSQC spectrum obtained at 600 MHz, D_2O, pH 7, 28°C for Lac-, 25°C for GM3-, 16°C for GM2-, 24°C for GM1a-, and 24°C GD3-FCHASE. The methyl resonance of internal acetone is at 31.07 ppm relative to external dioxane at 67.40 ppm. The error is ± 0.2 ppm for ^{13}C chemical shifts and ± 5°C for the sample temperature. The error is ± 0.8 ppm for 6a, 6b, 6d, 6e due to overlap. [b] A correction of +0.52 ppm was added to the

65

chemical shifts of the reference compounds (25, 27) to make them relative to dioxane set at 67.40 ppm. Differences of over 1 ppm between the chemical shifts of the FCHASE compound and the corresponding reference compound are indicated in bold. [c] C-3 and C-4 assignments have been reversed. [d] C-4 and C-6 assignments have been reversed.

Accurate mass determination of the enzymatic product of Cst-II from Lac-FCHASE indicated that two sialic acids had been added to the Lac-FCHASE acceptor (Figure 4). The proton resonances were assigned from COSY, 1D TOCSY and 1D NOESY and comparison of chemical shifts with known structures. The Glc H-1 to H-6 and Gal H-1 to H-4 resonances were assigned from 1D TOCSY on the H-1 resonances. The Neu5Ac resonances were assigned from COSY and confirmed by 1D NOESY. The 1D NOESY of the H-8, H-9 Neu5Ac resonances at 4.16 ppm was used to locate the H-9s and H-7 resonances (Michon *et al, supra.*). The singlet appearance of the H-7 resonance of Neu5Ac(2-3) arising from small vicinal coupling constants is typical of the 2-8 linkage (*Id.*). The other resonances were assigned from the HSQC spectrum and ^{13}C assignments for terminal sialic acid (*Id.*). The proton and ^{13}C carbon chemical shifts of the Gal unit were similar to those in GM3-FCHASE, indicating the presence of the αNeu5Ac(2-3)Gal linkage. The J_{HH} values, proton and ^{13}C chemical shifts of the two sialic acids were similar to those of αNeu5Ac(2-8)Neu5Ac in the α(2-8)-linked Neu5Ac trisaccharide (Salloway *et al.* (1996) *Infect. Immun.* **64**, 2945-2949) indicating the presence of that linkage. Hence, the product was identified as GD3-FCHASE. Sialylation at C-8 of Neu5Ac caused a downfield shift of -6.5 ppm in its C-8 resonance from 72.6 ppm to 79.1 ppm.

The inter-residue NOEs for GD3-FCHASE were also typical of the αNeu5Ac(2-8)αNeu5Ac(2-3)βGal sequence. The largest inter-residue NOEs from the two H-3$_{ax}$ resonances at 1.7 - 1.8 ppm of Neu5Ac(2-3) and Neu5Ac(2-8) are to the Gal H-3 and -8)Neu5Ac H-8 resonances. Smaller inter-residue NOEs to Gal H-4 and –8)Neu5Ac H-7 are also observed. NOEs on FCHASE resonances are also observed due the overlap of an FCHASE resonance with the H-3$_{ax}$ resonances (Gilbert *et al., supra.*). The inter-residue NOE from H-3$_{eq}$ of Neu5Ac(2-3) to Gal H-3 is also observed. Also, the intra-residues confirmed the proton assignments. The NOEs for the 2-8 linkage are the same as those observed for the -8Neu5Acα2- polysaccharide (Michon *et al., supra.*).

The sialic acid glycosidic linkages could also be confirmed by the use of the HMBC experiment which detects $^3J(C, H)$ correlations across the glycosidic bond. The results for both α-2,3 and α-2,8 linkages indicate the $^3J(C, H)$ correlations between the two Neu5Ac anomeric C-2 resonances and Gal H-3 and -8)Neu5Ac H-8 resonances. The intra-

5   residue correlations to the H-3_{ax} and H-3_{eq} resonances of the two Neu5Ac residues were also observed. The Glc (C-1, H-2) correlation is also observed since there was partial overlap of the crosspeaks at 101 ppm with the crosspeaks at 100.6 ppm in the HMBC spectrum.

Accurate mass determination of the enzymatic product of CgtA from GM3-FCHASE indicated that a *N*-acetylated hexose unit had been added to the GM3-FCHASE

10   acceptor (Figure 4). The product was identified as GM2-FCHASE since the glycoside proton and ^{13}C chemical shifts were similar to those for GM2 oligosaccharide (GM2OS) (Sabesan *et al.* (1984) *Can. J. Chem.* **62**, 1034-1045). From the HSQC spectrum for GM2-FCHASE and the integration of its proton spectrum, there are now two resonances at 4.17 ppm and 4.18 ppm along with a new anomeric "d1" and two NAc groups at 2.04 ppm. From TOCSY

15   and NOESY experiments, the resonance at 4.18 ppm was unambiguously assigned to Gal H-3 because of the strong NOE between H-1 and H-3. For βgalactopyranose, strong intra-residue NOEs between H-1 and H-3 and H-1 and H-5 are observed due to the axial position of the protons and their short interproton distances (Pavliak *et al.* (1993) *J. Biol. Chem.* **268**, 14146-14152; Brisson *et al.* (1997) *Biochemistry* **36**, 3278-3292; Sabesan *et al.* (1984) *Can.*

20   *J. Chem.* **62**, 1034-1045). From the TOCSY spectrum and comparison of the H1 chemical shifts of GM2-FCHASE and GM2OS (Sabesan *et al.*, *supra.*) the resonance at 4.17 ppm is assigned as Gal H-4. Similarly, from TOCSY and NOESY spectra, the H-1 to H-5 of GalNAc and Glc, and H-3 to H-6 of Neu5Ac were assigned. Due to broad lines, the multiplet pattern of the resonances could not be observed. The other resonances were

25   assigned from comparison with the HSQC spectrum of the precursor and ^{13}C assignments for GM2OS (Sabesan *et al.*, *supra.*). By comparing the HSQC spectra for GM3- and GM2-FCHASE glycosides, a -9.9 ppm downfield shift between the precursor and the product occurred on the Gal C-4 resonance. Along with intra-residue NOEs to H-3 and H-5 of βGalNAc, the inter-residue NOE from GalNAc H-1 to Gal H-4 at 4.17 ppm was also

30   observed confirming the βGalNAc(1-4)Gal sequence. The observed NOEs were those

expected from the conformational properties of the GM2 ganglioside (Sabesan *et al.*, *supra.*).

Accurate mass determination of the enzymatic product of CgtB from GM2-FCHASE indicated that a hexose unit had been added to the GM2-FCHASE acceptor

5   (Figure 4). The product was identified as GM1a-FCHASE since the glycoside ^{13}C chemical shifts were similar to those for the GM1a oligosaccharide (*Id.*). The proton resonances were assigned from COSY, 1D TOCSY and 1D NOESY. From a 1D TOCSY on the additional "e1" resonance of the product, four resonances with a mutltiplet pattern typical of β-galactopyranose were observed. From a 1D TOCSY and 1D NOESY on the H-1 resonances

10 of βGalNAc, the H-1 to H-5 resonances were assigned. The βGalNAc H-1 to H-4 multiplet pattern was typical of the β-galactopyranosyl configuration, confirming the identity of this sugar for GM2-FCHASE. It was clear that upon glylcosidation, the major perturbations occurred for the βGalNAc resonances, and there was –9.1 ppm downfield shift between the acceptor and the product on the GalNAc C-3 resonance. Also, along with intra-residue

15 NOEs to H-3, H-5 of Gal, an inter-residue NOE from Gal H-1 to GalNAc H-3 and a smaller one to GalNAc H-4 were observed, confirming the βGal(1-3)GalNAc sequence. The observed NOEs were those expected from the conformational properties of the GM1a ganglioside (Sabesan *et al.*, *supra.*).

There was some discrepancy with the assignment of the C-3 and C-4 βGal(1-

20 4) resonances in GM2OS and GM1OS which are reversed from the published data (Sabesan *et al.*, *supra.*). Previously, the assignments were based on comparison of ^{13}C chemical shifts with known compounds. For GM1a-FCHASE, the assignment for H-3 of Gal(1-4) was confirmed by observing its large vicinal coupling, $J_{2,3}$ = 10Hz, directly in the HSQC spectrum processed with 2 Hz / point in the proton dimension. The H-4 multiplet is much

25 narrower (< 5 Hz) due to the equatorial position of H-4 in galactose (Sabesan *et al.*, *supra.*). In Table 6, the C-4 and C-6 assignments of one of the sialic acids in (-8Neu5Ac2-)3 also had to be reversed (Michon *et al.*, *supra.*) as confirmed from the assignments of H-4 and H-6.

The ^{13}C chemical shifts of the FCHASE glycosides obtained from HSQC spectra were in excellent agreement with those of the reference oligosaccharides shown in

30 Table 6. Differences of over 1 ppm were observed for some resonances and these are due to

different aglycons at the reducing end. Excluding these resonances, the averages of the differences in chemical shifts between the FCHASE glycosides and their reference compound were less than ± 0.2 ppm. Hence, comparison of proton chemical shifts, J_{HH} values and ^{13}C chemical shifts with known structures, and use of NOEs or HMBC were all

5   used to determine the linkage specificity for various glycosyltransferases. The advantage of using HSQC spectra is that the proton assignment can be verified independently to confirm the assignment of the ^{13}C resonances of the atoms at the linkage site. In terms of sensitivity, the proton NOEs are the most sensitive, followed by HSQC then HMBC. Using a nano-NMR probe instead of a 5 mm NMR probe on the same amount of material reduced

10   considerably the total acquisition time, making possible the acquisition of an HMBC experiment overnight.

Discussion

In order to clone the LOS glycosyltransferases from *C. jejuni*, we employed an activity screening strategy similar to that which we previously used to clone the α-2,3-

15   sialyltransferase from *Neisseria meningitidis* (Gilbert *et al.*, *supra.*). The activity screening strategy yielded two clones which encoded two versions of the same α-2,3-sialyltransferase gene (*cst-I*). ORF analysis suggested that a 430 residue polypeptide is responsible for the α-2,3-sialyltransferase activity. To identify other genes involved in LOS biosynthesis, we compared a LOS biosynthesis locus in the complete genome sequence of *C. jejuni* NCTC

20   11168 to the corresponding locus from *C. jejuni* OH4384. Complete open reading frames were identified and analyzed. Several of the open reading frames were expressed individually in *E. coli*, including a β-1,4-*N*-acetylgalactosaminyl-transferase (*cgtA*), a β-1,3-galactosyltransferase (*cgtB*) and a bifunctional sialyltransferase (*cst-II*).

The *in vitro* synthesis of fluorescent derivatives of nanomole amounts of

25   ganglioside mimics and their NMR analysis confirm unequivocally the linkage specificity of the four cloned glycosyltransferases. Based on these data, we suggest that the pathway described in Figure 4 is used by *C. jejuni* OH4384 to synthesize a GT1a mimic. This role for *cgtA* is further supported by the fact that *C. jejuni* OH4342, which carries an inactive version of this gene, does not have β-1,4-GalNAc in its LOS outer core (Figure 1). The *cst-II* gene

30   from *C. jejuni* OH4384 exhibited both α-2,3- and α-2,8-sialyltransferase in an *in vitro* assay

while *cst-II* from *C. jejuni* O:19 (serostrain) showed only α-2,3-sialyltransferase activity (Table 5). This is consistent with a role for *cst-II* in the addition of a terminal α-2,8-linked sialic acid in *C. jejuni* OH4382 and OH4384, both of which have identical *cst-II* genes, but not in *C. jejuni* O:19 (serostrain, see Figure 1). There are 8 amino acid differences between the Cst-II homologues from *C. jejuni* O:19 (serostrain) and OH4382/84.

The bifunctionality of *cst-II* might have an impact on the outcome of the *C. jejuni* infection since it has been suggested that the expression of the terminal di-sialylated epitope might be involved in the development of neuropathic complications such as the Guillain-Barré syndrome (Salloway *et al.* (1996) *Infect. Immun.* **64**, 2945-2949). It is also worth noting that its bifunctional activity is novel among the sialyltransferases described so far. However, a bifunctional glycosyltransferase activity has been described for the 3-deoxy-D-manno-octulosonic acid transferase from *E. coli* (Belunis, C. J., and Raetz, C. R. (1992) *J. Biol. Chem.* **267**, 9988-9997).

The mono/bi-functional activity of *cst-II* and the activation/inactivation of *cgtA* seem to be two forms of phase variation mechanisms that allow *C. jejuni* to make different surface carbohydrates that are presented to the host. In addition to those small gene alterations that are found among the three O:19 strains (serostrain, OH4382 and OH4384), there are major genetic rearrangements when the loci are compared between *C. jejuni* OH4384 and NCTC 11168 (an O:2 strain). Except for the *prfB* gene, the *cst-I* locus (including *cysN* and *cysD*) is found only in *C. jejuni* OH4384. There are significant differences in the organization of the LOS biosynthesis locus between strains OH4384 and NCTC 11168. Some of the genes are well conserved, some of them are poorly conserved while others are unique to one or the other strain. Two genes that are present as separate ORFs (#5a: cgtA and #10a: NeuA) in OH4384 are found as an in-frame fusion ORF in NCTC 11168 (ORF #5b/#10b). β-*N*-acetylgalactosaminyltransferase activity was detected in this strain, which suggests that at least the *cgtA* part of the fusion may be active.

In summary, this Example describes the identification of several open reading frames that encode enzymes involved in the synthesis of lipooligosaccharides in *Campylobacter*.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent

5   applications cited herein are hereby incorporated by reference for all purposes.